

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À  
TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE  
APPLIQUÉES

PAR  
RACHELLE BELLERIVE

DISTRIBUTIONS MÉTA-ELLIPTIQUES : INFÉRENCE  
STATISTIQUE ET APPLICATIONS

NOVEMBRE 2012

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

## REMERCIEMENTS

Je tiens à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire. Tout d'abord, j'aimerais exprimer toute ma gratitude envers M. Jean-François Quessy, professeur au Département des mathématiques et d'informatique de l'Université du Québec à Trois-Rivières, sans qui je n'aurais pu compléter ce mémoire. Je le remercie sincèrement de m'avoir proposé ce projet passionnant et pour sa confiance, son encouragement et sa grande disponibilité. Son aide m'a été précieuse et j'ai bien apprécié travailler avec lui.

Je remercie également M. François Meunier, professeur au Département des mathématiques et d'informatique de l'Université du Québec à Trois-Rivières, et M. Ali Assani, professeur au Département des sciences humaines de l'Université du Québec à Trois-Rivières, d'avoir accepté d'évaluer ce mémoire. Leurs commentaires ont certainement aidés à améliorer sa qualité.

Je désire également remercier mon amoureux, Alex, pour avoir été là pour moi. Il a toujours su m'encourager et m'aider tout au long de mes études, et surtout dans les moments les plus difficiles. Merci pour ton écoute et pour ta bonne humeur. Je remercie aussi mes frères et ma soeur, Matthieu, Jody et Thomas, pour tous les rires et les bons moments qu'ils m'ont fait vivre. Merci d'être comme vous êtes.

Mes études à la maîtrise ont été financées en majeure partie par des subventions de recherche octroyées à M. Jean-François Quessy par le *Fonds Québécois de Recherche Nature et Technologies* et par le *Conseil National*

*de Recherche en Sciences Naturelles et en Génie* du Canada. Je remercie également l'*Institut des Sciences Mathématiques* du Québec, ainsi que le département de Mathématiques et d'Informatique de l'Université du Québec à Trois-Rivières pour les bourses d'étude qu'ils m'ont accordées.

# Table des matières

<b>Remerciements</b>	<b>ii</b>
<b>Liste des tableaux</b>	<b>viii</b>
<b>Liste des figures</b>	<b>x</b>
<b>Chapitre 1. Introduction</b>	<b>1</b>
<b>Chapitre 2. Les lois méta-elliptiques</b>	<b>4</b>
2.1 Distributions elliptiques multivariées . . . . .	4
2.2 Description de quelques lois elliptiques . . . . .	12
2.3 Simulation de vecteurs de loi elliptique . . . . .	14
2.4 Théorème de Sklar et copules . . . . .	14
2.5 Les copules elliptiques . . . . .	20
<b>Chapitre 3. Revue des méthodes d'inférence pour les copules</b>	<b>22</b>
3.1 Estimation de paramètres . . . . .	22
3.1.1 Inversion du tau de Kendall . . . . .	23
3.1.2 Inversion du rho de Spearman . . . . .	27
3.1.3 Maximum de la pseudo-vraisemblance . . . . .	28
3.2 Sélection d'une famille de copules via les tests d'adéquation . .	29
3.2.1 Processus d'adéquation de la copule empirique . . . . .	30

3.2.2	Processus d'adéquation de Kendall . . . . .	31
3.2.3	Processus d'adéquation de Spearman . . . . .	32
<b>Chapitre 4. Nouveaux tests d'adéquation de lois méta-elliptiques : aperçu du Chapitre 5</b>		<b>34</b>
4.1	Estimation de la matrice de corrélation . . . . .	34
4.2	Copies de la variable aléatoire $\mathcal{G}^2$ . . . . .	35
4.3	Une statistique de Cramér-von Mises . . . . .	36
4.4	Approximation des $p$ -valeurs . . . . .	38
<b>Chapitre 5. Statistical Procedures for the Selection of a Multi- dimensional Meta-elliptical Distribution</b>		<b>39</b>
5.1	Introduction . . . . .	40
5.2	The meta-elliptical family of dependence functions . . . . .	44
5.2.1	Elliptically contoured distributions . . . . .	44
5.2.2	Elliptical copulas and meta-elliptical models . . . . .	48
5.3	Goodness-of-fit procedure in the case of a fixed generator . . . . .	50
5.3.1	Context . . . . .	50
5.3.2	Estimation of the correlation matrix . . . . .	51
5.3.3	Test statistic and parametric bootstrap . . . . .	53
5.3.4	Investigation of the size and power of the test . . . . .	57
5.4	Goodness-of-fit procedure in the case of a parametric generator . . . . .	59
5.4.1	A minimum-distance method . . . . .	59
5.4.2	Investigation of the power of the test . . . . .	61
5.4.3	Performance of the minimum-distance estimator . . . . .	62
5.5	Illustrations on real data . . . . .	63
5.5.1	The Danish fire insurance data . . . . .	63

5.5.2	The Upper Mississippi river data . . . . .	65
5.5.3	The Oil currency data . . . . .	67
5.5.4	The Cook & Johnson data set revisited . . . . .	68
<b>Chapitre 6. Conclusion</b>		<b>82</b>
<b>Bibliographie</b>		<b>85</b>
<b>Annexe A. Programmes en Matlab</b>		<b>91</b>
A.1	Simulation de variables aléatoires . . . . .	91
A.1.1	Quelques lois univariées . . . . .	91
A.1.2	Quelques lois multivariées . . . . .	92
A.2	Procédures utilitaires . . . . .	93
A.2.1	Estimation de la matrice de corrélation . . . . .	93
A.2.2	Densité de la loi Pearson type II . . . . .	93
A.2.3	Fonction de répartition Pearson II inverse . . . . .	93
A.3	Adéquation à une loi méta-elliptique de générateur fixe . . . . .	94
A.3.1	Fonction de répartition de la variable aléatoire $\mathcal{G}^2$ . . . . .	94
A.3.2	Calcul de la statistique de test $V_n$ . . . . .	95
A.3.3	Exécution du test basé sur $V_n$ . . . . .	96
A.4	Adéquation à une loi méta-elliptique de générateur paramétrique . . . . .	97
A.4.1	Calcul de la statistique de test $W_n$ . . . . .	97
A.4.2	Exécution du test basé sur $W_n$ . . . . .	98
A.5	Adéquation à une loi univariée . . . . .	99
A.5.1	Fonction de répartition de quelques lois . . . . .	99
A.5.2	Calcul de la statistique de test $S_n$ . . . . .	99
A.5.3	Exécution du test basé sur $S_n$ . . . . .	100
A.6	Approximation de la puissance des tests . . . . .	101

A.6.1 Test basé sur $V_n$ . . . . .	101
A.6.2 Test basé sur $W_n$ . . . . .	104



## LISTE DES TABLEAUX

2.1	Quelques distributions elliptiques et leurs générateurs . . . . .	13
5.1	Estimation, based on 1 000 replicates, of $n$ times the mean-squared error $\widehat{\text{MSE}}_R(R_n)$ of $R_n$ for estimating the correlation matrix $R$ in the equicorrelated case . . . . .	74
5.2	Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for bivariate meta-elliptical models when $n = 100$ . . . . .	75
5.3	Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for bivariate meta-elliptical models when $n = 250$ . . . . .	76
5.4	Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for meta-elliptical models in the equicorrelated case for $d \in \{3, 4\}$ and $n = 100$ . . . . .	77
5.5	Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for the parametric meta-elliptical models $T$ and $\text{Pe}$ in the equicorrelated case with $\rho = 1/2$ for $d \in \{2, 3\}$ when $n \in \{50, 100\}$ . . . . .	78
5.6	Estimation, based on 1 000 replicates, of $n$ times the mean-squared errors of $\hat{\theta}_n$ for estimating the parameter of the elliptical generator in the Student and Pearson type II copulas when $n = 50$ . . . . .	79

5.7	Results of the goodness-of-fit tests on the Danish fire insurance and Upper Mississippi river data sets for the null hypothesis of a meta-elliptical distribution with a fixed generator ( $p$ -values and critical values estimated from $M = 1\,000$ parametric bootstrap samples) . . . . .	80
5.8	Results of the goodness-of-fit tests on the Oil currency and Uranium exploration data sets for the null hypothesis of a meta-elliptical distribution with a fixed generator ( $p$ -values and critical values estimated from $M = 250$ parametric bootstrap samples) . . . . .	81

## LISTE DES FIGURES

5.1	Three-dimensional histograms of the realizations of 5 000 pairs from the $T_1$ (upper left panel), $T_3$ (upper right panel), $T_9$ (middle left panel), Normal (middle right panel), $Pe_1$ (lower left panel) and $Pe_5$ (lower right panel) copulas when $\tau = .5$ .	70
5.2	Scatterplot (upper left panel) and density plot (upper right panel) of the normalized ranks and histograms (lower panels) for the Danish insurance data set.	71
5.3	Probability that both claims in the Danish fire insurance data set exceed $x_0$ (in millions of Euros).	72
5.4	Plot of $\Psi_n$ (dashed lines) and $\Psi_{G^2}$ for the $N$ (upper left panel), $T_6$ (upper right panel), $T_{16}$ (lower left panel) and $Pe_2$ (lower right panel) elliptical copulas for the Oil currency data set.	73

# CHAPITRE 1

## INTRODUCTION

Soit  $\mathbf{X} = (X_1, \dots, X_d)^\top$ , un vecteur aléatoire dans  $\mathbb{R}^d$ . Un aspect statistique important et intéressant est de modéliser la structure de dépendance de ce vecteur. À cette fin, une approche utile et populaire consiste à avoir recours à la théorie des copules. L'usage des copules a connu un essor fulgurant ces dernières années, autant du côté des avancées théoriques que des applications dans divers domaines tels la finance, l'hydrologie et la climatologie. Le point de départ de cette théorie est le Théorème de Sklar (1959). Ce résultat stipule que si  $H$  est une fonction de répartition continue dont les marges sont  $F_1, \dots, F_d$ , alors il existe une unique copule  $C : [0, 1]^2 \rightarrow [0, 1]$  telle que  $H(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}$ . Puisque toute l'information sur les marges est contenue dans  $F_1, \dots, F_d$ , il en découle que  $C$  rassemble toute l'information sur la dépendance entre les composantes de  $\mathbf{X}$ .

La modélisation de la dépendance s'effectue généralement en tentant d'*ajuster* un modèle théorique à une série d'observations. On a donc recours à des *familles* de modèles. Une de celles-ci est l'intéressante classe des copules elliptiques, qui sont extraites des lois elliptiques *classiques*. Cette classe est

très riche, comprenant entre autres la fameuse loi Normale, ainsi que la loi de Student multidimensionnelle. Outre ces distributions, la famille elliptique comprend une panoplie d'autres modèles, la rendant très flexible pour la modélisation de diverses structures stochastiques pour un vecteur  $\mathbf{X}$ . Les copules elliptiques héritent de cette propriété enviables.

En général, le choix d'un modèle approprié s'effectue à l'aide d'un test d'adéquation. Il s'agit d'appliquer le test à un certain nombre de modèles et de conserver ceux pour lesquels le test conclut à un non-rejet. Avec un peu de chance, un seul de ces modèles est accepté ! Cette idée de sélection de modèles est particulièrement importante lorsqu'on travaille avec la famille elliptique, vu le grand choix de structures possibles. Quelques tests d'adéquation ont été développés assez récemment, dont ceux basés sur la copule empirique (voir Genest et al., 2009) et sur le processus de Kendall (voir Genest et al., 2006). Ces tests peuvent, en principe, être utilisés pour les modèles elliptiques. Toutefois, ils sont assez difficile à mettre en œuvre, ce qui appelle à des procédures alternatives mieux adaptées.

L'objectif de ce mémoire est de construire de nouveaux tests d'adéquation pour les copules elliptiques. L'idée centrale est basée sur la soi-disante *partie radiale* qui caractérise une loi elliptique. Des estimateurs semi-paramétriques de cette variable aléatoire sont d'abord proposés ; ils sont basés sur les rangs des observations, ce qui les rend invariant sous des transformations monotones croissantes, donc indépendants des lois marginales. Leur distribution empirique est ensuite comparée à la loi théorique attendue sous l'hypothèse nulle via une statistique de Cramér-von Mises. Une approche plus générale

qui inclut les cas où la partie radiale possède un paramètre inconnu est également explorée en détails. La méthode d'estimation de ce paramètre ainsi que la statistique d'adéquation associée sont basées sur une approche à *distance minimale*.

Le mémoire est organisé comme suit. Le Chapitre 2 fait office d'introduction aux principaux résultats théoriques relatifs aux lois elliptiques et aux copules; on y fait en outre le lien entre ces deux notions, ce qui donne naissance aux modèles méta-elliptiques. Au Chapitre 3, on passe en revue quelques méthodes *classiques* d'estimation de paramètres et d'adéquation pour les modèles de copules. Les chapitres 4-5 forment le cœur de ce travail. En fait, le Chapitre 5 est constitué de l'intégral d'un article scientifique soumis à un numéro spécial sur les copules chapeauté par la *Revue de la Société Française de Statistique*. Cet article intitulé *Statistical Procedures for the Selection of a Multidimensional Meta-elliptical Distribution* développe en détails de nouvelles méthodes d'adéquation pour les modèles méta-elliptiques et étudie leurs propriétés. Des illustrations détaillées concernant les jeux de données *Danish Fire Insurance*, *Upper Mississippi River*, *Oil Currency* et *Uranium exploration* sont aussi présentées. Le code des programmes en langage Matlab qui ont été utilisés est offert à l'Annexe A.

## CHAPITRE 2

### LES LOIS MÉTA-ELLIPTIQUES

#### 2.1 Distributions elliptiques multivariées

Soit  $\mathbf{X} = (X_1, \dots, X_d)^\top$ , un vecteur aléatoire dans  $\mathbb{R}^d$ . On dit que  $\mathbf{X}$  suit une loi elliptique si sa fonction caractéristique  $\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}^\top \mathbf{X}})$ , où  $i = \sqrt{-1}$ , est de la forme

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp(i\mathbf{t}^\top \mu) \Upsilon\left(\frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t}\right), \quad (2.1)$$

où  $\mathbf{t} = (t_1, \dots, t_d)^\top$  est le vecteur transposé de  $\mathbf{t}_{1 \times d} = (t_1, \dots, t_d)$ ,  $\mu = (\mu_1, \dots, \mu_d)$  est un vecteur de moyenne,  $\Sigma \in \mathbb{R}^{d \times d}$  est une matrice de covariance et  $\Upsilon : \mathbb{R}^+ \rightarrow \mathbb{R}$  est tel que

$$\Upsilon\left(\sum_{j=1}^d t_j^2\right)$$

est une fonction caractéristique à  $d$  dimensions. On appelle  $\Upsilon$  le générateur caractéristique.

Une caractérisation utile des vecteurs de loi elliptique provient du Corollaire 1 de Fang & Zhang (1990). Ce résultat stipule que  $\mathbf{X}$  est de loi elliptique si et seulement si

$$\mathbf{X} \stackrel{d}{=} \mu + \mathcal{G} A \mathcal{U}, \quad (2.2)$$

où  $A \in \mathbb{R}^{d \times d}$  est une matrice telle que  $A A^\top = \Sigma$  et  $\mathcal{U} = (U_1, \dots, U_d)^\top$  est un vecteur aléatoire uniformément distribué sur la sphère de rayon 1, c'est-à-dire que  $U_j \in [-1, 1]$  pour chaque  $j \in \{1, \dots, d\}$  et  $\|\mathcal{U}\|^2 = \mathcal{U}^\top \mathcal{U} = 1$ . De plus,  $\mathcal{G}$  est une variable aléatoire positive, indépendante de  $\mathcal{U}$ , dont la densité est

$$\psi_{\mathcal{G}}(\gamma) = \frac{2 \pi^{d/2}}{\Gamma(d/2)} \gamma^{d-1} g(\gamma^2/2), \quad \gamma \geq 0. \quad (2.3)$$

Ici,  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  est un générateur tel que

$$\int_0^\infty \gamma^{d-1} g(\gamma^2/2) d\gamma = \frac{\Gamma(d/2)}{(2\pi)^{d/2}}. \quad (2.4)$$

Cette condition assure que  $\psi_{\mathcal{G}}$  est une densité. On écrit alors  $\mathbf{X} \sim \mathcal{E}(\mu, \Sigma, g)$ .

**Remarque 2.1.** *Comme cas particulier de la représentation 2.2, on retrouve la caractérisation des distributions sphériques lorsque  $\mu = 0$  et  $\Sigma = I_n$ , où  $I_n$  est la matrice identité. On a alors  $\mathbf{X} \stackrel{d}{=} \mathcal{G} \mathcal{U}$ .*

Différents choix de fonction  $g$  génèrent différents modèles elliptiques. Par exemple, lorsque

$$g(\gamma) = \frac{1}{(2\pi)^{d/2}} e^{-\gamma},$$

la loi de  $\mathbf{X}$  est la Normale  $d$ -dimensionnelle. Dans ce cas, la densité de  $\mathcal{G}$  est

$$\psi_{\mathcal{G}}(\gamma) = \frac{\gamma^{d-1} e^{-\gamma^2/2}}{\Gamma(\frac{d}{2}) 2^{d/2-1}}.$$



De là, on peut montrer que la densité de  $\mathcal{G}^2$  est donnée par

$$\psi_{\mathcal{G}^2}(\gamma) = \frac{\gamma^{d/2-1} e^{-\gamma/2}}{\Gamma\left(\frac{d}{2}\right) 2^{d/2}}.$$

Autrement dit,  $\mathcal{G}^2$  suit une loi khi-carré à  $d$  degrés de liberté. La loi Normale fut l'une des raisons pour laquelle on s'est intéressé aux distributions elliptiques. On verra plus loin que plusieurs caractéristiques de cette loi sont partagées par les modèles de la famille elliptique. Par exemple, on montrera un résultat sur les combinaisons linéaires de variables aléatoires elliptiques. D'autres modèles, tels la Student multivariée, seront décrits à la section 2.2.

La densité d'une variable aléatoire de loi  $\mathcal{E}(\mu, \Sigma, g)$ , si elle existe, s'écrit

$$h_{\mu, \Sigma, g}(\mathbf{x}) = |\Sigma|^{-1/2} g\left\{\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right\}, \quad (2.5)$$

où  $\mathbf{x} = (x_1, \dots, x_d)$ . De plus, les densités marginales de  $(X_j - \mu_j)/\sigma_j$ , où  $\mu_j = E(X_j)$  et  $\sigma_j^2 = \text{var}(X_j)$ , sont toutes identiques; elles sont de la forme

$$q_g(x) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_{x^2}^{\infty} (y - x^2)^{\frac{d-3}{2}} g(y/2) dy. \quad (2.6)$$

Ces dernières font partie de la famille des lois univariées elliptiques. On peut montrer que la fonction de répartition associée à  $q_g$  est

$$Q_g(x) = \int_{-\infty}^x q_g(s) ds = \frac{1}{2} + \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^x \int_{s^2}^{\infty} (y - u^2)^{\frac{d-3}{2}} g(y/2) dy ds. \quad (2.7)$$

Pour montrer que  $\int_{-\infty}^{\infty} q_g(x) dx = 1$ , c'est-à-dire que  $q_g$  est bel et bien une densité, on note d'abord qu'avec le changement de variable  $z = y - x^2$ , on a

$$q_g(x) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^{\infty} z^{\frac{d-3}{2}} g\{(z + x^2)/2\} dz.$$

En posant  $t = \sqrt{z}$ , on obtient ensuite

$$q_g(x) = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^\infty t^{d-2} g\{(t^2 + x^2)/2\} dt.$$

On a alors

$$\int_{-\infty}^\infty q_g(x) = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^\infty \int_{-\infty}^\infty t^{d-2} g\{(t^2 + x^2)/2\} dx dt.$$

En utilisant les coordonnées polaires  $t = r \sin \theta$  et  $x = r \cos \theta$ , et en faisant appel à la propriété (2.4), on a

$$\begin{aligned} \int_{-\infty}^\infty q_g(x) &= \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^\pi \int_0^\infty (r \sin \theta)^{d-2} g\left(\frac{r^2 \sin^2 \theta + r^2 \cos^2 \theta}{2}\right) r dr d\theta \\ &= \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^\pi (\sin \theta)^{d-2} \left\{ \int_0^\infty r^{d-1} g\left(\frac{r^2}{2}\right) dr \right\} d\theta \\ &= \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^\pi (\sin \theta)^{d-2} \times \frac{\Gamma(d/2)}{2\pi^{d/2}} d\theta \\ &= \frac{\Gamma(d/2)}{\sqrt{\pi} \Gamma(\frac{d-1}{2})} \int_0^\pi (\sin \theta)^{d-2} d\theta. \end{aligned}$$

Il reste à montrer que pour tout  $d \geq 2$ ,

$$\int_0^\pi (\sin \theta)^{d-2} d\theta = \frac{\sqrt{\pi} \Gamma(\frac{d-1}{2})}{\Gamma(d/2)}.$$

D'abord, l'égalité est vraie pour  $d = 2$  car

$$\int_0^\pi (\sin \theta)^0 d\theta = \int_0^\pi d\theta = \pi,$$

alors que de l'autre côté, on a

$$\frac{\sqrt{\pi} \Gamma(\frac{2-1}{2})}{\Gamma(\frac{2}{2})} = \frac{\sqrt{\pi} \Gamma(1/2)}{\Gamma(1)} = \frac{\sqrt{\pi} \cdot \sqrt{\pi}}{1} = \pi.$$

Elle est aussi vraie pour  $d = 3$  car

$$\int_0^\pi \sin \theta d\theta = -\cos \theta \Big|_0^\pi = -(-1) - (-1) = 2,$$

alors que

$$\frac{\sqrt{\pi} \Gamma\left(\frac{3-1}{2}\right)}{\Gamma\left(\frac{3}{2}\right)} = \frac{\sqrt{\pi} \Gamma(1)}{\frac{1}{2} \Gamma\left(\frac{1}{2}\right)} = \frac{\sqrt{\pi}}{\frac{1}{2} \sqrt{\pi}} = 2.$$

Supposons maintenant que l'égalité est vraie pour un certain  $d \geq 2$ . En utilisant l'identité trigonométrique

$$\int_0^\pi (\sin \theta)^d d\theta = \left(\frac{d-1}{d}\right) \int_0^\pi (\sin \theta)^{d-2} d\theta,$$

on obtient

$$\begin{aligned} \int_0^\pi (\sin \theta)^d d\theta &= \left(\frac{d-1}{d}\right) \frac{\sqrt{\pi} \Gamma\left(\frac{d-1}{2}\right)}{\Gamma(d/2)} \\ &= \sqrt{\pi} \frac{(d-1)/2 \Gamma\left(\frac{d-1}{2}\right)}{(d/2) \Gamma(d/2)} \\ &= \sqrt{\pi} \frac{\Gamma\left(\frac{d-1}{2} + 1\right)}{\Gamma\left(\frac{d}{2} + 1\right)} \\ &= \frac{\sqrt{\pi} \Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d+2}{2}\right)}. \end{aligned}$$

Autrement dit, si l'égalité est vraie pour un certain  $d \geq 2$ , alors elle est vraie pour  $d+2$ . Comme elle est vraie pour  $d=2$  et  $d=3$ , il s'ensuit par induction qu'elle est vraie pour tout  $d \geq 2$ . Par conséquent,  $q_g(x)$  est une densité.

Le résultat suivant sera utilisé plus loin pour construire des tests d'adéquation pour les modèles de dépendance extraits des lois elliptiques.

**Proposition 2.1.** *Si  $\mathbf{X} \sim \mathcal{E}(\mu, \Sigma, g)$ , alors la loi de*

$$\tilde{\mathbf{X}} = (\mathbf{X} - \mu)^\top \Sigma^{-1} (\mathbf{X} - \mu)$$

*est la même que la loi de  $\mathcal{G}^2$  dont la densité est donnée par*

$$\psi_{\mathcal{G}^2}(\gamma) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \gamma^{d/2-1} g(\gamma/2). \quad (2.8)$$

**Démonstration.** En utilisant la représentation  $\mathbf{X} \stackrel{d}{=} \mu + \mathcal{G} A \mathcal{U}$  et du fait que  $\Sigma = A A^\top$ , on obtient

$$\begin{aligned}
 \tilde{\mathbf{X}} &= (\mathbf{X} - \mu)^\top \Sigma^{-1} (\mathbf{X} - \mu) \\
 &= (\mu + \mathcal{G} A \mathcal{U} - \mu)^\top \Sigma^{-1} (\mu + \mathcal{G} A \mathcal{U} - \mu) \\
 &= \mathcal{G} (A \mathcal{U})^\top (A A^\top)^{-1} \mathcal{G} (A \mathcal{U}) \\
 &= \mathcal{G}^2 \mathcal{U}^\top A^\top (A^\top)^{-1} A^{-1} A \mathcal{U} \\
 &= \mathcal{G}^2 \mathcal{U}^\top \mathcal{U} \\
 &= \mathcal{G}^2,
 \end{aligned}$$

où le fait que  $\|\mathcal{U}\|^2 = \mathcal{U}^\top \mathcal{U} = 1$  permet d'obtenir la dernière égalité. Pour finir, la loi de  $\mathcal{G}^2$  s'obtient à partir de la densité de  $\mathcal{G}$  donnée à l'équation (2.3).

Il s'agit de poser  $\mathcal{G}^2 = g(\mathcal{G})$ , où  $g(x) = x^2$ . On a alors

$$g^{-1}(\gamma) = \sqrt{\gamma} \quad \text{et} \quad \frac{d}{d\gamma} g^{-1}(\gamma) = \frac{1}{2\sqrt{\gamma}}.$$

Ainsi,

$$\psi_{\mathcal{G}^2}(\gamma) = \psi_{\mathcal{G}}(\sqrt{\gamma}) \times \frac{1}{2\sqrt{\gamma}} = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \gamma^{d/2-1} g(\gamma),$$

ce qui complète la démonstration.  $\diamond$

On peut montrer que la loi d'une combinaison linéaire de variables aléatoires de loi elliptique est aussi elliptique.

**Proposition 2.2.** Si  $\mathbf{X} \sim \mathcal{E}(\mu, \Sigma, g)$ , alors

$$\mathbf{X}^* = b + B \mathbf{X} \sim \mathcal{E}(b + B\mu, B\Sigma B^\top, g).$$

où  $B \in \mathbb{R}^{d \times d}$  et  $b \in \mathbb{R}$ .

**Démonstration.** De la représentation  $\mathbf{X} \stackrel{d}{=} \mu + \mathcal{G} A \mathcal{U}$ , on peut écrire

$$\mathbf{X}^* = b + B \mathbf{X} \stackrel{d}{=} b + B(\mu + \mathcal{G} A \mathcal{U}) \stackrel{d}{=} b + B\mu + \mathcal{G} B A \mathcal{U}.$$

Ainsi, on a la représentation  $\mathbf{X}^* = \mu^* + \mathcal{G} A^* \mathcal{U}$ , où  $\mu^* = b + B\mu$  et  $A^* = B A$ .

Finalement, comme  $AA^T = \Sigma$ , il ne reste qu'à observer que

$$(BA)(BA)^T = (BA)(A^T B^T) = B \Sigma B^T,$$

ce qui complète la preuve.  $\diamond$

On peut aussi montrer que la combinaison linéaire de deux vecteurs aléatoires indépendants de loi elliptique suit également une loi elliptique.

**Proposition 2.3.** *Soient deux variables aléatoires indépendantes  $\mathbf{X}_1$  et  $\mathbf{X}_2$  telles que*

$$\mathbf{X}_1 \sim \mathcal{E}(\mu_1, \Sigma, g_1) \quad \text{et} \quad \mathbf{X}_2 \sim \mathcal{E}(\mu_2, \alpha \Sigma, g_2),$$

*pour un certain  $\alpha > 0$ . Alors, pour  $\gamma_1, \gamma_2 \in \mathbb{R}$ ,*

$$\gamma_1 \mathbf{X}_1 + \gamma_2 \mathbf{X}_2 \sim \mathcal{E}(\gamma_1 \mu_1 + \gamma_2 \mu_2, \Sigma, g_3),$$

*où*

$$g_3(s) = g_1\left(\frac{1}{2} \gamma_1^2 s\right) \times g_2\left(\frac{1}{2} \alpha \gamma_2^2 s\right).$$

**Démonstration.** Soient  $\phi_{\mathbf{X}_1}$  et  $\phi_{\mathbf{X}_2}$ , les fonctions caractéristiques de  $\mathbf{X}_1$  et  $\mathbf{X}_2$ , respectivement. Ainsi, de l'équation (2.1), on a

$$\phi_{\mathbf{X}_1}(\mathbf{t}) = e^{i\mathbf{t}^\top \mu_1} \Upsilon_1\left(\frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t}\right) \quad \text{et} \quad \phi_{\mathbf{X}_2}(\mathbf{t}) = e^{i\mathbf{t}^\top \mu_2} \Upsilon_2\left(\frac{1}{2} \alpha \mathbf{t}^\top \Sigma \mathbf{t}\right)$$

pour certaines fonctions  $\Upsilon_1 : \mathbb{R}^+ \rightarrow \mathbb{R}$  et  $\Upsilon_2 : \mathbb{R}^+ \rightarrow \mathbb{R}$ . Par conséquent, les fonctions caractéristiques de  $\mathbf{X}_1 - \mu_1$  et  $\mathbf{X}_2 - \mu_2$  sont respectivement

$$\phi_{\mathbf{X}_1 - \mu_1}(\mathbf{t}) = \Upsilon_1 \left( \frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t} \right) \quad \text{et} \quad \phi_{\mathbf{X}_2 - \mu_2}(\mathbf{t}) = \Upsilon_2 \left( \frac{1}{2} \alpha \mathbf{t}^\top \Sigma \mathbf{t} \right).$$

Ensuite, puisque la fonction caractéristique d'une somme de variables aléatoires indépendantes est le produit des fonctions caractéristiques individuelles,

$$\begin{aligned} \phi_{\gamma_1(\mathbf{X}_1 - \mu_1) + \gamma_2(\mathbf{X}_2 - \mu_2)}(\mathbf{t}) &= \phi_{\gamma_1(\mathbf{X}_1 - \mu_1)}(\mathbf{t}) \times \phi_{\gamma_2(\mathbf{X}_2 - \mu_2)}(\mathbf{t}) \\ &= \phi_{\mathbf{X}_1 - \mu_1}(\gamma_1 \mathbf{t}) \times \phi_{\mathbf{X}_2 - \mu_2}(\gamma_2 \mathbf{t}) \\ &= \Upsilon_1 \left( \frac{1}{2} (\gamma_1 \mathbf{t})^\top \Sigma (\gamma_1 \mathbf{t}) \right) \times \Upsilon_2 \left( \frac{1}{2} \alpha (\gamma_2 \mathbf{t})^\top \Sigma (\gamma_2 \mathbf{t}) \right) \\ &= \Upsilon_1 \left( \frac{1}{2} \gamma_1^2 \mathbf{t}^\top \Sigma \mathbf{t} \right) \times \Upsilon_2 \left( \frac{1}{2} \alpha \gamma_2^2 \mathbf{t}^\top \Sigma \mathbf{t} \right). \end{aligned}$$

Ainsi,

$$\phi_{\gamma_1(\mathbf{X}_1 - \mu_1) + \gamma_2(\mathbf{X}_2 - \mu_2)}(\mathbf{t}) = \Upsilon_3 \left( \frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t} \right),$$

où

$$\Upsilon_3(s) = \Upsilon_1(\gamma_1^2 s) \times \Upsilon_2(\alpha \gamma_2^2 s).$$

De là,

$$\begin{aligned} \phi_{\gamma_1 \mathbf{X}_1 + \gamma_2 \mathbf{X}_2}(\mathbf{t}) &= \phi_{\gamma_1(\mathbf{X}_1 - \mu_1) + \gamma_2(\mathbf{X}_2 - \mu_2) + \gamma_1 \mu_1 + \gamma_2 \mu_2}(\mathbf{t}) \\ &= e^{i\mathbf{t}(\gamma_1 \mu_1 + \gamma_2 \mu_2)} \phi_{\gamma_1(\mathbf{X}_1 - \mu_1) + \gamma_2(\mathbf{X}_2 - \mu_2)}(\mathbf{t}) \\ &= e^{i\mathbf{t}(\gamma_1 \mu_1 + \gamma_2 \mu_2)} \Upsilon_3 \left( \frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t} \right), \end{aligned}$$

ce qui termine la preuve. ◇

## 2.2 Description de quelques lois elliptiques

Au tableau 2.1, on trouve plusieurs exemples de lois elliptiques, incluant la loi Normale multivariée. Dans ce cas, on a déjà mentionné que si  $\mathbf{X}$  suit cette loi, alors la distribution de  $\mathcal{G}^2 = \tilde{\mathbf{X}}$ , où

$$\tilde{\mathbf{X}} = (\mathbf{X} - \mu)^\top \Sigma^{-1} (\mathbf{X} - \mu),$$

est une khi-carré à  $\nu = d$  degrés de liberté.

Dans le cas où  $\mathbf{X}$  est une variable aléatoire de loi de Pearson type II, alors

$$g_\theta(\gamma) = \frac{\Gamma\left(\frac{d}{2} + \theta + 1\right)}{\pi^{d/2} \Gamma(\theta + 1)} (1 - 2\gamma)^\theta, \quad \gamma \in [0, 1/2].$$

Comme cas particulier de l'Équation (2.8), on trouve

$$\psi_{\mathcal{G}^2}(\gamma) = \frac{\Gamma\left(\frac{d}{2} + \theta + 1\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma(\theta + 1)} \gamma^{d/2-1} (1 - \gamma)^\theta, \quad \gamma \in [0, 1].$$

Il s'agit de la densité de la loi Beta de paramètres  $\alpha = d/2$  et  $\beta = \theta + 1$ .

Pour la loi de Student à  $\nu$  degrés de liberté, on a

$$g_\nu(\gamma) = \frac{(\pi\nu)^{-d/2} \Gamma\left(\frac{d+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{2\gamma}{\nu}\right)^{-(\nu+d)/2},$$

ce qui fait que

$$\psi_{\mathcal{G}^2}(\gamma) = \frac{\nu^{\nu/2} \Gamma\left(\frac{d+\nu}{2}\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} \gamma^{d/2-1} (\gamma + \nu)^{-(\nu+d)/2}.$$

TABLE 2.1: Quelques distributions elliptiques et leurs générateurs

Loi elliptique	$g(\gamma)$	Conditions
Normale	$\frac{1}{(2\pi)^{d/2}} \exp(-\gamma)$	
Student	$\frac{(\pi\nu)^{-d/2} \Gamma\left(\frac{d+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{2\gamma}{\nu}\right)^{-(\nu+d)/2}$	
Exponentielle	$\frac{\theta_2 \Gamma\left(\frac{d}{2}\right) \theta_1^{d/2\theta_2}}{(2\pi)^{d/2} \Gamma\left(\frac{d}{2\theta_2}\right)} \exp(-\theta_1 \gamma^{\theta_2})$	$\theta_1, \theta_2 > 0$
Cauchy	$\pi^{-(d+1)/2} \Gamma\left(\frac{d+1}{2}\right) (1+2\gamma)^{-(d+1)/2}$	
Pearson type II	$\frac{\Gamma\left(\frac{d}{2} + \theta + 1\right)}{\pi^{d/2} \Gamma(\theta + 1)} (1-2\gamma)^\theta, \gamma \in [0, 1/2]$	$\theta > -1$
Kotz	$\frac{s \Gamma\left(\frac{d}{2}\right) r^{(2N+d-2)/(2s)} \gamma^{N-1} e^{-r\gamma^s}}{\pi^{(d-1)/2} \Gamma\left(\frac{2N+d-2}{2s}\right)}$	$r, s > 0, 2N+d > 2$



## 2.3 Simulation de vecteurs de loi elliptique

Afin de générer des vecteurs aléatoires provenant d'une loi elliptique, il est utile d'exploiter la représentation stochastique (2.2). Ceci induit l'algorithme de simulation décrit dans la suite. Ici, on souhaite générer  $\mathbf{X}$  de loi  $\mathcal{E}(\mu, \Sigma, g)$ .

1. Trouver la matrice  $A$  telle que  $A^\top A = \Sigma$ , c'est-à-dire la décomposition de Choleski de  $\Sigma$  ;
2. Simuler un vecteur  $\mathbf{U} = (U_1, \dots, U_d)$  uniformément distribué sur la sphère de rayon 1 :
  - (a) Générer un vecteur aléatoire  $\mathbf{Z} = (Z_1, \dots, Z_d)$ , où  $Z_1, \dots, Z_d$  sont indépendantes et de loi  $\mathbb{N}(0, 1)$  ;
  - (b) Prendre

$$\mathbf{U} = \frac{\mathbf{Z}}{\|\mathbf{Z}\|_2} = \left( \frac{Z_1}{\sqrt{Z_1^2 + \dots + Z_d^2}}, \dots, \frac{Z_d}{\sqrt{Z_1^2 + \dots + Z_d^2}} \right) ;$$

3. Générer  $\mathcal{G}$ , une variable aléatoire caractérisée par le générateur  $g$  ;
4. Poser  $\mathbf{X} = \mu + \mathcal{G}A\mathbf{U}$ .

## 2.4 Théorème de Sklar et copules

Depuis quelques années, les copules ont beaucoup gagnée en popularité. On les utilise fréquemment pour étudier la dépendance d'un couple de variables aléatoires  $(X_1, X_2)$ . Les copules ont l'avantage d'être extrêmement flexibles

pour modéliser la dépendance et sont relativement faciles à manipuler. Sur-tout, une copule ne dépend pas des lois marginales.

Mathématiquement parlant, une copule à deux dimensions est une fonction définie sur  $[0, 1]^2$ . Ainsi,  $C : [0, 1]^2 \rightarrow [0, 1]$  est une copule si elle satisfait les deux propriétés suivantes :

- (i) pour tout  $u \in [0, 1]$ ,  $C(u, 0) = C(0, u) = 0$  et  $C(u, 1) = C(1, u) = u$  ;
- (ii) pour tout  $0 \leq u_1 \leq u_2 \leq 1$  et  $0 \leq v_1 \leq v_2 \leq 1$ ,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Le théorème de Sklar (1959) fait le lien entre la définition de copule et la loi bivariée d'un couple de variables aléatoires  $(X_1, X_2)$ . Il s'agit du résultat le plus important pour l'étude des copules.

**Théorème 2.1.** *Soit  $H$ , la fonction de répartition d'un couple aléatoire  $(X_1, X_2)$ . Si  $F_1$  et  $F_2$  sont les marges de  $H$ , alors il existe une copule  $C$  telle que*

$$H(x_1, x_2) = C \{F_1(x_1), F_2(x_2)\}. \quad (2.9)$$

*Si  $F_1$  et  $F_2$  sont continues, alors  $C$  est unique. On dit alors que  $C$  est la copule de  $H$ .*

À noter que  $C$  correspond à la loi conjointe de  $U_1 = F_1(X_1)$  et  $U_2 = F_2(X_2)$ ,

où  $U_1$  et  $U_2$  sont de loi uniforme sur  $(0, 1)$ . En effet,

$$\begin{aligned}
 P(U_1 \leq u_1, U_2 \leq u_2) &= P\{F_1(X_1) \leq u_1, F_2(X_2) \leq u_2\} \\
 &= P\{X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2)\} \\
 &= H\{F_1^{-1}(u_1), F_2^{-1}(u_2)\} \\
 &= C\{F_1 \circ F_1^{-1}(u_1), F_2 \circ F_2^{-1}(u_2)\} \\
 &= C(u_1, u_2).
 \end{aligned}$$

Dans cette chaîne d'égalités, on constate que

$$C(u_1, u_2) = H\{F_1^{-1}(u_1), F_2^{-1}(u_2)\}. \quad (2.10)$$

Il s'agit d'une façon d'extraire la copule d'une loi  $H$ . Par exemple, cela permet de montrer que la copule associée à l'indépendance bivariée est  $C(u_1, u_2) = u_1 u_2$ . En effet, puisque l'indépendance entre  $(X_1, X_2)$  est caractérisée par

$$H(x_1, x_2) = F_1(x_1) \times F_2(x_2),$$

on obtient par une application de l'équation (2.10) que

$$C(u_1, u_2) = F_1\{F_1^{-1}(u_1)\} \times F_2\{F_2^{-1}(u_2)\} = u_1 u_2.$$

La dépendance positive parfaite survient lorsque les variables  $U_1 = F_1(X_1)$  et  $U_2 = F_2(X_2)$  sont telles que  $U_1 = U_2$ . Dans ce cas, la copule associée est

$$\begin{aligned}
 M(u_1, u_2) &= P(U_1 \leq u_1, U_1 \leq u_2) \\
 &= P\{U_1 \leq \min(u_1, u_2)\} \\
 &= \min(u_1, u_2).
 \end{aligned}$$

On a la dépendance négative parfaite quand  $U_2 = 1 - U_1$ . Dans ce cas, la copule associée est

$$\begin{aligned} W(u_1, u_2) &= P(U_1 \leq u_1, 1 - U_1 \leq u_2) \\ &= P(U_1 \leq u_1, U_1 \geq 1 - u_2) \\ &= P(1 - u_2 \leq U_1 \leq u_1) \\ &= \max(u_1 + u_2 - 1, 0). \end{aligned}$$

Les fonctions  $W$  et  $M$ , qui sont des copules, s'appellent les bornes inférieure et supérieure de Fréchet–Hoeffding. On peut montrer que toute copule  $C$  est telle que pour tout  $(u_1, u_2) \in [0, 1]^2$ ,

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2).$$

Le théorème de Sklar (1959) se généralise au cas à  $d$  dimensions. Ainsi, si  $H$  est la fonction de répartition de  $(X_1, \dots, X_d)$  dont les marges  $F_1, \dots, F_d$  sont continues, alors il existe une unique fonction  $C : [0, 1]^d \rightarrow [0, 1]$  telle que

$$H(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}.$$

Il existe une version du théorème de Sklar (1959) pour les fonctions de survie. Pour la décrire, soit la fonction de survie de  $(X_1, X_2)$  définie par

$$\bar{H}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2).$$

On définit, de la même façon, les survies marginales par

$$\bar{F}_1(x_1) = P(X_1 > x_1) = 1 - F_1(x_1) \quad \text{et} \quad \bar{F}_2(x_2) = P(X_2 > x_2) = 1 - F_2(x_2).$$

Alors la version *fonctions de survie* du théorème de Sklar stipule qu'il existe une unique fonction  $C : [0, 1]^2 \rightarrow [0, 1]$  telle que

$$\bar{H}(x_1, x_2) = \hat{C}\{\bar{F}_1(x_1), \bar{F}_2(x_2)\}. \quad (2.11)$$

Ici,  $\widehat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$  est appelée la copule de survie.

Pour extraire la copule d'une certaine fonction de répartition  $d$ -dimensionnelle  $H$  dont les marges sont

$$F_j(x_j) = \lim_{x_i \rightarrow \infty, i \neq j} H(x_1, \dots, x_d),$$

on remarque que de la représentation (2.11), on a

$$C(u_1, \dots, u_d) = H \{ F_1^{-1}(u_1), \dots, F_d^{-1}(u_d) \}. \quad (2.12)$$

**Exemple 2.1.** Soit la fonction de répartition bivariable

$$H(x_1, x_2) = \sqrt{\frac{x_1 x_2}{x_1 + x_2 - x_1 x_2}},$$

où  $(x_1, x_2) \in [0, 1]^2$ . D'abord, les marges de  $H$  sont

$$F_1(x_1) = \lim_{x_2 \rightarrow \infty} H(x_1, x_2) = H(x_1, 1) = \sqrt{x_1}$$

et

$$F_2(x_2) = \lim_{x_1 \rightarrow \infty} H(x_1, x_2) = \sqrt{x_2}.$$

Comme  $F_1^{-1}(u_1) = u_1^2$  et  $F_2^{-1}(u_2) = u_2^2$ , on déduit de l'équation (2.12) dans le cas où  $d = 2$  que

$$C(u_1, u_2) = H(u_1^2, u_2^2) = \sqrt{\frac{u_1^2 u_2^2}{u_1^2 + u_2^2 - u_1^2 u_2^2}} = (u_1^{-2} + u_2^{-2} - 1)^{-1/2}.$$

Cette copule est un cas particulier du modèle de Clayton lorsque  $\theta = 2$ .

La copule de Clayton est l'une des nombreuses copules appartenant à la famille des copules Archimédiennes. Ces copules sont de la forme

$$C_\phi(u_1, u_2) = \phi^{-1} \{ \phi(u_1) + \phi(u_2) \}, \quad (2.13)$$

où  $\phi : [0, 1] \rightarrow [0, \infty)$  est une fonction continue, strictement décroissante et convexe tel que  $\phi(1) = 0$ . Cette fonction s'appelle le générateur de  $C_\phi$ . Pour la copule de Clayton, le générateur est

$$\phi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}.$$

De l'équation (2.13), la copule de Clayton est donc de la forme

$$C_\theta(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}. \quad (2.14)$$

La famille Archimédienne est particulièrement intéressante pour la forme simple de ses membres. Elle contient également une grande variété de modèles, incluant les copules de Clayton, Frank et Gumbel-Hougaard.

Parmi les propriétés des copules Archimédiennes ce trouve la symétrie de  $C_\phi$ . C'est-à-dire que pour tout  $(u_1, u_2) \in [0, 1]^2$ ,

$$C_\phi(u_1, u_2) = C_\phi(u_2, u_1).$$

De plus,  $C_\phi$  est associative, c'est-à-dire que

$$C_\phi \{u_1, C_\phi(u_2, u_3)\} = C_\phi \{C_\phi(u_1, u_2), u_3\}$$

pour tout  $(u_1, u_2, u_3) \in [0, 1]^3$ .

Il existe de nombreux autres modèles de copules, notamment les modèles à valeurs extrêmes. Ces derniers permettent de modéliser des phénomènes extrêmes tels les ouragans ou la crue des eaux. La prochaine section est dédiée aux copules *meta-elliptiques* que l'on extrait des lois elliptiques multivariées ; elle est centrale pour la suite de ce mémoire.

## 2.5 Les copules elliptiques

On dit que  $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$  suit une loi méta-elliptique si et seulement si sa copule associée est elliptique. Autrement dit, si  $F_1, \dots, F_d$  sont les marges de  $\mathbf{Y}$  (pas nécessairement elliptiques), alors

$$(Q_g^{-1} \circ F_1(Y_1), \dots, Q_g^{-1} \circ F_d(Y_d)) \sim \mathcal{E}(0, \Sigma, g), \quad (2.15)$$

où  $Q_g$  est la fonction de répartition donnée par l'équation (2.7). Comme cas particulier,  $\mathbf{Y}$  suit une loi elliptique *classique*  $\mathcal{E}(0, \Sigma, g)$  lorsque  $F_j = Q_g$  pour tout  $j \in \{1, \dots, d\}$ .

**Exemple 2.2.** Soit  $\mathbf{Y} = (Y_1, \dots, Y_d)$  de marges  $F_1, \dots, F_d$  et de copule de Student à  $\nu$  degrés de libertés. Alors

$$(t_\nu^{-1} \circ F_1(Y_1), \dots, t_\nu^{-1} \circ F_d(Y_d))$$

suit une loi de Student  $d$ -dimensionnelle à  $\nu$  degrés de liberté, où  $t_\nu$  est la fonction de répartition de la loi Student univariée à  $\nu$  degrés de liberté.

De l'équation (2.5), la fonction de répartition associée à la loi elliptique  $\mathcal{E}(0, \Sigma, g)$  est de la forme

$$H_{\Sigma, g}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} |\Sigma|^{-1/2} g\left(\frac{1}{2} \mathbf{s}^\top \Sigma^{-1} \mathbf{s}\right) d s_1 \cdots d s_d,$$

où  $\mathbf{s} = (s_1, \dots, s_d)$ . De (2.12), on déduit que la copule associée à  $H_{\Sigma, g}$  est

$$C_{\Sigma, g}(u_1, \dots, u_d) = H_{\Sigma, g}\{Q_g^{-1}(u_1), \dots, Q_g^{-1}(u_d)\}.$$

La copule elliptique de  $H_{\Sigma, g}$  a donc la forme implicite

$$C_{\Sigma, g}(u_1, \dots, u_d) = \int_{-\infty}^{Q_g^{-1}(u_1)} \cdots \int_{-\infty}^{Q_g^{-1}(u_d)} |\Sigma|^{-1/2} g\left(\frac{1}{2} \mathbf{s}^\top \Sigma^{-1} \mathbf{s}\right) d s_1 \cdots d s_d.$$

Par exemple, pour la loi Normale  $d$ -dimensionnelle, on a

$$C(u_1, \dots, u_d) = \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{d/2} \Sigma^{1/2}} \exp\left(-\frac{\mathbf{s}^\top \Sigma^{-1} \mathbf{s}}{2}\right) d\mathbf{s} dt.$$

Soit maintenant  $\mathbf{X} = (X_1, \dots, X_d)$ , un vecteur de loi  $\mathcal{E}(0, \Sigma, g)$ . Dans ce cas, on sait que les marges de  $\mathbf{X}$  sont  $Q_g$ . Ainsi, la loi de

$$\mathcal{U} = (Q_g(X_1), \dots, Q_g(X_d))$$

est la copule  $C_{\Sigma, g}$ . Ensuite, si on pose

$$\mathbf{Y} = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)),$$

alors  $\mathbf{Y}$  a une structure de dépendance méta-elliptique  $C_{\Sigma, g}$ , et les marges  $F_1, \dots, F_d$ . En résumé, la copule de

$$\mathbf{Y} = (F_1^{-1} \circ Q_g(X_1), \dots, F_d^{-1} \circ Q_g(X_d)) \quad (2.16)$$

est  $C_{\Sigma, g}$  et ses marges sont  $F_1, \dots, F_d$ .

L'équation (2.16) fournit une façon de simuler des observations d'une loi méta-elliptique  $\mathcal{ME}(\Sigma, g)$ . Il s'agit d'abord de simuler un vecteur  $\mathbf{X}$  de loi  $\mathcal{E}(\mathbf{0}, \Sigma, g)$  avec la méthode décrite à la section 2.3. Ensuite, avec le choix désiré des marges  $F_1, \dots, F_d$ , on définit  $\mathbf{Y}$  comme à la représentation (2.16).



## CHAPITRE 3

### REVUE DES MÉTHODES D'INFÉRENCE POUR LES COPULES

Dans ce chapitre, quelques techniques d'inférence pour les modèles de copules seront décrites. Celles-ci ont été développées, en grande partie, dans les dernières années. Pour simplifier les explications, la description des méthodes d'estimation se fera pour le cas à deux variables seulement. L'extension au cas multidimensionnel est assez directe, mais ne sera pas explorée ici.

#### 3.1 Estimation de paramètres

Supposons que l'on observe des paires  $(X_1, Y_1), \dots, (X_n, Y_n)$  de variables aléatoires. Ces dernières peuvent être vues comme des copies d'un vecteur aléatoire  $(X, Y)$ . Afin de modéliser la structure de dépendance de  $(X, Y)$ , on peut supposer que le modèle de copule sous-jacent appartient à une certaine famille paramétrique. Autrement dit, on suppose que la copule est de la forme  $C_\theta$ , où  $\theta \in \Theta$  est un paramètre inconnu. Il est donc nécessaire de l'estimer. Trois méthodes populaires sont décrites dans la suite, à savoir l'inversion du

tau de Kendall, l'inversion du rho de Spearman, et la méthode du maximum de la pseudo-vraisemblance.

### 3.1.1 Inversion du tau de Kendall

Le tau de Kendall mesure la dépendance entre deux variables aléatoires ; sa magnitude ne dépend pas des lois marginales. Tout comme le coefficient de corrélation *classique*, la valeur du tau de Kendall se situe entre  $-1$  et  $1$ . La valeur  $1$  correspond à la *dépendance positive parfaite*, alors que  $-1$  signale une *dépendance négative parfaite*.

Pour décrire formellement le tau de Kendall, on doit définir les notions de concordance et de discordance. Soient donc deux vecteurs indépendants  $(X_1, Y_1)$  et  $(X_2, Y_2)$  de même loi  $H$ . On dit qu'ils sont concordants, noté  $\mathcal{C}$ , si

$$X_1 < X_2, Y_1 < Y_2 \quad \text{ou si} \quad X_1 > X_2, Y_1 > Y_2.$$

Sinon, on dit qu'ils sont discordants, ce qui est noté  $\mathcal{D}$ . Le tau de Kendall associé à une loi  $H$  est alors défini par

$$\tau(H) = P(\mathcal{C}) - P(\mathcal{D}).$$

Dès lors, on constate que  $\tau(H) = 1$  lorsque  $(X, Y) \sim H$  sont parfaitement concordants, alors que  $\tau(H) = -1$  si  $(X, Y)$  sont parfaitement discordants.

Bien qu'en apparence, la définition du tau de Kendall fait intervenir la loi  $H$  d'un couple  $(X, Y)$ , on peut montrer que  $\tau(H)$  ne dépend pas de ses marges  $F$  et  $G$ , mais seulement de la copule  $C$  telle que  $H(x, y) = C\{F(x), G(y)\}$ .

C'est l'objet du résultat suivant.

**Proposition 3.1.** *Soit  $H$ , une fonction de répartition bivariée de marges continues. Alors*

$$\tau(H) = \tau(C) = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1,$$

où  $C$  est l'unique copule associée à  $H$ .

**Démonstration.** Soient les paires indépendantes  $(X_1, Y_1)$  et  $(X_2, Y_2)$  de loi  $H$ . Comme  $P(\mathcal{D}) = 1 - P(\mathcal{C})$ , on a par la définition du tau de Kendall que

$$\begin{aligned} \tau(H) &= P(\mathcal{C}) - P(\mathcal{D}) \\ &= 2P(\mathcal{C}) - 1 \\ &= 2P(X_1 < X_2, Y_1 < Y_2) + 2P(X_1 > X_2, Y_1 > Y_2) - 1. \end{aligned}$$

Puisque  $(X_1, Y_1)$  et  $(X_2, Y_2)$  sont de même loi, les deux probabilités précédentes sont identiques, ce qui permet d'écrire

$$\tau(H) = 4P(X_1 < X_2, Y_1 < Y_2) - 1.$$

Ensuite, en conditionnant par rapport à  $(X_2, Y_2)$ , on obtient

$$\begin{aligned} P(X_1 < X_2, Y_1 < Y_2) &= \int_{\mathbb{R}^2} P(X_1 < x, Y_1 < y) dH(x, y) \\ &= \int_{\mathbb{R}^2} H(x, y) dH(x, y) \\ &= \int_{\mathbb{R}^2} C\{F(x), G(y)\} dC\{F(x), G(y)\}, \end{aligned}$$

où la dernière égalité est une conséquence du théorème de Sklar. En effectuant le changement de variables  $u = F(x)$  et  $v = G(y)$ , on déduit que

$$P(X_1 < X_2, Y_1 < Y_2) = \int_{[0,1]^2} C(u, v) dC(u, v),$$

ce qui complète la preuve.  $\diamond$

Dans la suite, on décrira une version empirique du tau de Kendall basée sur des observations  $(X_1, Y_1), \dots, (X_n, Y_n)$  indépendantes et identiquement distribuées selon une loi  $H$ . Le tau de Kendall empirique est alors défini par

$$\tau_n = \frac{C_n - D_n}{\binom{n}{2}},$$

où  $C_n$  et  $D_n$  sont respectivement le nombre de concordances et de discordances parmi les  $n(n-1)/2$  paires de couples possibles.

Reprenons maintenant le contexte d'une famille paramétrique de copules qui s'écrivent  $C_\theta$ , où  $\theta$  est un paramètre univarié. Dans ce cas, le tau de Kendall dépend du paramètre de telle sorte que

$$\tau(C_\theta) = k(\theta) = 4 \int_{[0,1]^2} C_\theta(u, v) dC_\theta(u, v) - 1.$$

La méthode d'inversion du tau de Kendall consiste à faire coïncider la valeur du tau de Kendall théorique  $k(\theta)$  avec celle de sa version empirique  $\tau_n$ . Cette idée est dans le même esprit que la méthode des moments classique. Ainsi, l'estimateur de  $\theta$  est la solution  $\theta_n$  de l'équation

$$k(\theta_n) = \tau_n,$$

c'est-à-dire

$$\theta_n = k^{-1}(\tau_n).$$

Parfois, une formule explicite est disponible pour  $k^{-1}$ . C'est le cas pour la copule Normale ainsi que pour les modèles de Clayton et de Farlie–Gumbel–Morgenstern décrits dans les trois exemples suivants.

**Exemple 3.1.** Soit  $h_\rho$ , la densité Normale à deux variables de moyennes nulles, de variances unitaires et de corrélation  $\rho$ . La fonction de répartition associée est

$$H_\rho(x, y) = \int_{-\infty}^x \int_{-\infty}^y h_\rho(s, t) dt ds.$$

Les lois marginales de  $H_\rho$  sont

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds.$$

Par une application de l'équation (2.10), on déduit que la copule Normale s'écrit implicitement

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} h_\rho(s, t) dt ds.$$

Un résultat de Lindskog et al. (2003) indique que le tau de Kendall associé à  $C_\rho$  est

$$\tau(C_\rho) = \frac{2}{\pi} \sin^{-1}(\rho).$$

Ainsi, la méthode d'inversion du tau de Kendall amène l'estimateur

$$\rho_n = \sin\left(\frac{\pi}{2} \tau_n\right).$$

**Exemple 3.2.** Pour la copule de Clayton décrite à l'équation (2.14), on montre que

$$k(\theta) = \frac{\theta}{\theta + 2}.$$

Ainsi,

$$k^{-1}(t) = \frac{2t}{1-t},$$

ce qui fait que l'estimateur de  $\theta$  est

$$\theta_n = \frac{2\tau_n}{1-\tau_n}.$$

**Exemple 3.3.** La copule de Farlie–Gumbel–Morgenstern s'écrit sous la forme

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v), \quad \text{où } \theta \in [-1, 1].$$

On montre alors que  $k(\theta) = 2\theta/9$ , ce qui fait que  $k^{-1}(t) = 9t/2$ . Donc, on estimerait  $\theta$  avec  $\theta_n = 9\tau_n/2$ .

À noter que lorsque la fonction  $k$  n'est pas inversible, on peut avoir recours à des méthodes numériques, tel la méthode de Newton, pour approximer la solution  $\theta$  de  $k(\theta) - \tau_n = 0$ .

### 3.1.2 Inversion du rho de Spearman

Plutôt que d'utiliser le tau de Kendall, on pourrait estimer un paramètre univarié  $\theta$  en inversant le rho de Spearman. À l'instar du tau de Kendall, cette mesure de dépendance ne dépend que de la copule  $C$  sous-jacente à une loi bivariée  $H$ . On peut la définir par

$$\rho_S(H) = \rho_S(C) = 12 \int_{[0,1]^2} C(u, v) du dv - 3.$$

Une version empirique de  $\rho_S$  est donnée par le coefficient de corrélation de rangs, à savoir

$$\rho_n = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - S_i)^2,$$

où  $R_i$  est le rang de  $X_i$  parmi  $X_1, \dots, X_n$  et  $S_i$  est le rang de  $Y_i$  parmi  $Y_1, \dots, Y_n$ . La méthode d'inversion du rho de Spearman consiste donc à trouver la valeur  $\theta_n$  telle que  $\gamma(\theta) = \rho_n$ , où

$$\gamma(\theta) = \rho_S(C_\theta) = 12 \int_{[0,1]^2} C_\theta(u, v) du dv - 3.$$

Pour la copule de Farlie–Gumbel–Morgenstern présentée à l'exemple 3.3, on a  $\gamma(\theta) = \theta/3$ . Par conséquent, l'estimateur de  $\theta$  par l'inversion du rho de Spearman est simplement  $\theta_n = 3\rho_n$ .

### 3.1.3 Maximum de la pseudo-vraisemblance

Sur la base d'observations  $(X_1, Y_1), \dots, (X_n, Y_n)$  i.i.d.  $H_\theta$ , où  $\theta \in \Theta$ , la fonction de log-vraisemblance est donnée par

$$L(\theta) = \sum_{i=1}^n \ln h_\theta(X_i, Y_i), \quad (3.1)$$

où

$$h_\theta(x, y) = \frac{\partial^2}{\partial x \partial y} H_\theta(x, y)$$

est la densité de  $H_\theta$ . Supposons maintenant que

$$H_\theta(x, y) = C_\theta \{F_1(x), F_2(y)\},$$

c'est-à-dire que le paramètre  $\theta$  à estimer est associé uniquement à la structure de dépendance, et non aux marges. On montre alors que

$$h_\theta(x, y) = c_\theta \{F_1(x), F_2(y)\} \times dF_1(x) \times dF_2(y),$$

où  $c_\theta(u, v) = \partial^2 C_\theta(u, v) / \partial u \partial v$  est la densité de  $C_\theta$ . Une autre façon d'écrire la log-vraisemblance (3.1) est alors

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \ln c_\theta \{F_1(X_i), F_2(Y_i)\} \times dF_1(X_i) \times dF_2(Y_i) \\ &= \sum_{i=1}^n \ln c_\theta \{F_1(X_i), F_2(Y_i)\} + \sum_{i=1}^n \ln dF_1(X_i) + \sum_{i=1}^n \ln dF_2(Y_i). \end{aligned}$$

L'estimateur du maximum de vraisemblance consiste à choisir  $\hat{\theta}_{MV} \in \Theta$  tel que  $L(\theta)$  est maximisée. Puisque les deux derniers termes de l'équation précédente ne dépendent pas de  $\theta$ , il s'agit donc de maximiser

$$L(\theta) = \sum_{i=1}^n \ln c_{\theta} \{F_1(X_i), F_2(Y_i)\}.$$

Cependant, les marges  $F_1$  et  $F_2$  sont généralement inconnues. L'idée de Genest et al. (1995) consiste à simplement estimer les lois marginales par les fonctions de répartition empiriques

$$F_{n1}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x) \quad \text{et} \quad F_{n2}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_i \leq y). \quad (3.2)$$

Une autre version de  $L(\theta)$  est alors

$$\tilde{L}(\theta) = \sum_{i=1}^n \ln c_{\theta} \{F_{n1}(X_i), F_{n2}(Y_i)\},$$

que l'on appelle la *pseudo log-vraisemblance*. L'estimateur de  $\theta$  est alors le maximum de  $\tilde{L}(\theta)$ . Il est entièrement basé sur les rangs des observations.

### 3.2 Sélection d'une famille de copules via les tests d'adéquation

Soit un vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_d)^{\top}$ . On s'intéresse à la copule modélisant la dépendance entre les variables marginales  $X_1, \dots, X_d$  de  $\mathbf{X}$ . Il est alors naturel de se demander à quelle famille appartient cette copule  $C$ . Pour les copules en général, il existe trois principaux tests d'adéquations sur une copule soient, le test basé sur le processus de copule empirique, celui



basé sur le processus de Kendall et celui basé sur le processus de Spearman.

On confrontera les hypothèses

$$H_0 : C \in \{C_\theta; \theta \in \Theta\} \quad \text{et} \quad H_1 : C \notin \{C_\theta; \theta \in \Theta\}.$$

### 3.2.1 Processus d'adéquation de la copule empirique

L'estimation d'une copule inconnue se fait généralement via la copule empirique définie par

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1} \{F_{n1}(X_i) \leq u, F_{n2}(Y_i) \leq v\},$$

où les marges empiriques  $F_{n1}$  et  $F_{n2}$  sont décrites à l'équation (3.2). Une façon de mesurer l'écart entre l'information sur la dépendance provenant des données et celle fournie par le modèle sous  $\mathcal{H}_0$  consiste à considérer la statistique de Cramér-von Mises

$$V_n = n \int_0^1 \int_0^1 \{C_n(u, v) - C_{\theta_n}(u, v)\}^2 du dv,$$

où  $\theta_n$  est un estimateur convergent pour  $\theta$ . Cette idée a été proposée par Quessy (2005) et étudiée plus en détails par Genest et al. (2009) et Berg & Quessy (2009). Puisque  $V_n$  est en général assez difficile à calculer, une version asymptotiquement équivalente a été proposée par Genest et al. (2009), à savoir

$$V_{n,N} = n \int_0^1 \int_0^1 \{C_n(u, v) - C_N(u, v)\}^2 du dv,$$

où  $C_N(u, v)$  est la copule empirique d'observations *artificiellement* simulées de la loi  $C_{\theta_n}$ . La validité de cette astuce a été rigoureusement démontrée par

Genest & Rémillard (2008). Afin d'obtenir des  $p$ -valeurs pour un test basé sur  $V_{n,N}$ , la technique du *bootstrap* paramétrique est généralement employée. Il s'agit de générer  $M$  échantillons indépendants de taille  $n$  à partir de la loi  $C_{\theta_n}$ . On calcule alors la statistique de test pour chacun de ces échantillons *bootstrap*, ce qui amène  $V_{n,N}^{(1)}, \dots, V_{n,N}^{(M)}$ . La  $p$ -valeur du test est alors

$$\hat{p} = \frac{1}{M} \sum_{h=1}^M \mathbf{1} \left( V_{n,N}^{(h)} > V_{n,N} \right).$$

On rejette l'hypothèse  $\mathcal{H}_0$  lorsque  $\hat{p} < \alpha$  pour un certain seuil  $\alpha$  fixé.

### 3.2.2 Processus d'adéquation de Kendall

La transformation intégrale de probabilité associée à une copule  $C$  est définie par  $K(t) = P \{C(U, V) \leq t\}$ , où  $(U, V) \sim C$ . Il a été démontré par Genest & Rivest (2001) que

$$K(t) = t + \int_0^1 C_{01} \{C_v^{-1}(t), v\} dv,$$

où

$$C_{01}(u, v) = \frac{\partial}{\partial v} C(u, v) \quad \text{et} \quad C_v^{-1}(t) = \{u | C(u, v) = t\}.$$

Lorsque la copule appartient à un modèle paramétrique, on écrit  $K_\theta$ . Une façon d'estimer  $K_\theta$  consiste à utiliser l'estimateur

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\xi_{i,n} \leq t),$$

où

$$\xi_{i,n} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{1}(X_j < X_i, Y_j < Y_i).$$

Il a d'abord été proposé par Genest & Rivest (1993) et son comportement asymptotique a été étudié par Barbe et al. (1996). En supposant que  $K_\theta$  possède une forme explicite, on pourrait utiliser la statistique de test

$$W_n = n \int_0^1 \{K_n(t) - K_{\theta_n}(t)\}^2 dK_\theta(t).$$

À l'instar de la statistique  $V_{n,N}$ , la  $p$ -valeur d'un test basé sur  $W_n$  sera obtenue par l'application du *bootstrap* paramétrique. À noter que si  $K_\theta$  n'est pas explicite, alors on peut utiliser une version  $W_{n,N}$  de  $W_n$  dans laquelle  $K_{\theta_n}$  est remplacée par une estimation  $K_N$ .

### 3.2.3 Processus d'adéquation de Spearman

Un test basé sur le processus d'adéquation de Spearman fut développé par Mesfioui et al. (2009). Ce test est similaire à la procédure basée sur la statistique  $W_n$ . Ces auteurs définissent d'abord la fonction de dépendance de Spearman d'un couple  $(X_1, X_2)$  de loi  $H$  par

$$S(t) = P \{F_1(X_1) \times F_2(X_2) \leq t\},$$

où  $F_1$  et  $F_2$  sont les lois marginales. Cette fonction tire son nom du fait que

$$\rho_S(H) = 12 \int_0^1 \{1 - S(t)\} dt - 3,$$

où  $\rho_S(H)$  est le rho de Spearman de  $H$ . La fonction de dépendance de Spearman s'exprime en fonction de la copule  $C$  de  $H$  via la formule

$$S(t) = t + \int_t^1 C_{01}\left(\frac{t}{w}, w\right) dw,$$

où  $C_{01}(u, v) = \partial C(u, v) / \partial v$ . Soient maintenant  $R_i$ , le rang de  $X_{i1}$  parmi  $X_{11}, \dots, X_{n1}$  et  $S_i$ , le rang de  $X_{i2}$  parmi  $X_{12}, \dots, X_{n2}$ . La version empirique de  $S$  est alors donnée par

$$S_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\varphi_{i,n} \leq t),$$

où

$$\varphi_{i,n} = F_{n1}(X_{i1}) \times F_{n2}(X_{i2}) = \frac{1}{n^2} R_i S_i.$$

Si  $S_\theta$ ,  $\theta \in \Theta$ , est associée à une famille paramétrique de copules, alors une statistique d'adéquation est

$$W_n^* = n \int_0^1 \{S_n(t) - S_{\theta_n}(t)\}^2 dS_{\theta_n}(t).$$

On peut montrer que

$$W_n^* = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{1 - S_{\theta_n}(\varphi_{i,n} \vee \varphi_{j,n})\} - \sum_{i=1}^n \{1 - S_{\theta_n}^2(\varphi_{i,n})\} + \frac{n}{3},$$

où  $a \vee b = \max(a, b)$ . La  $p$ -valeur s'obtient du *bootstrap* paramétrique.

## CHAPITRE 4

### NOUVEAUX TESTS D'ADÉQUATION DE LOIS MÉTA-ELLIPTIQUES : APERÇU DU CHAPITRE 5

Ce chapitre se veut une brève introduction à l'article scientifique qui constitue l'intégrale du Chapitre 5. Cette article a été soumis pour publication pour la *Special issue on copula modelling* chapeautée par la *Revue de la Société Française de Statistique*.

Ce papier introduit de nouvelles méthodes d'adéquation pour les modèles de copules elliptiques. Les statistiques de test proposées sont très flexibles puisqu'elles permettent que le générateur elliptique soit paramétrique. Au passage, les méthodes d'estimation des différents paramètres sont étudiées en détails à l'aide de simulations Monte-Carlo.

#### 4.1 Estimation de la matrice de corrélation

On a vu au Chapitre 2 que les copules elliptiques possèdent deux composantes, à savoir une matrice de corrélation  $R$  et un générateur  $g$ . Pour esti-

mer les éléments de la matrice  $R$ , on utilise un résultat de Lindskog et al. (2003) qui stipule que le tau de Kendall de la paire  $(X_i, X_j)$  du vecteur  $\mathbf{X} = (X_1, \dots, X_d)$  est donné par

$$\tau_{ij} = \frac{2}{\pi} \sin^{-1}(R_{ij}).$$

Ainsi, si  $\tau_{n,ij}$  est le tau de Kendall empirique associé à la paire  $(X_i, X_j)$ , alors l'estimateur des moments de  $R_{ij}$  est

$$R_{n,ij} = \sin\left(\frac{\pi}{2} \tau_{n,ij}\right).$$

L'estimation de la matrice  $R$  est donc fournie par la matrice  $R_n$  dont les éléments hors-diagonaux sont  $(R_n)_{ij} = R_{n,ij}$ , alors que  $(R_n)_{ii} = 1$ .

## 4.2 Copies de la variable aléatoire $\mathcal{G}^2$

Soit une variable aléatoire  $\mathbf{Y} = (Y_1, \dots, Y_d)$  de loi méta-elliptique  $\mathcal{ME}(R, g)$  de marges  $F_1, \dots, F_d$ . On a vu à la Section 2.5 que la variable aléatoire

$$\mathbf{X} = (Q_g^{-1} \circ F_1(Y_1), \dots, Q_g^{-1} \circ F_d(Y_d))$$

suit une loi elliptique  $\mathcal{E}(\mathbf{0}, R, g)$ . De plus, la Proposition 2.1 indique que  $\mathcal{G}^2 = \mathbf{X}^\top R^{-1} \mathbf{X}$  suit une loi dont la densité est donnée à l'équation (2.8).

Afin de vérifier l'hypothèse que des observations proviennent d'une certaine loi méta-elliptique  $\mathcal{ME}(R, g)$ , on construira d'abord des *copies* de la variable aléatoire  $\mathcal{G}^2$ . Premièrement, on pose pour chaque  $i \in \{1, \dots, n\}$

$$\mathbf{X}_{i,n} = (Q_g^{-1} \circ F_{n1}(Y_{i1}), \dots, Q_g^{-1} \circ F_{nd}(Y_{id}))^\top,$$

où  $F_{n1}, \dots, F_{nd}$  sont les fonctions de répartition empiriques univariées. Ensuite, on définit

$$\mathcal{G}_{i,n}^2 = \mathbf{X}_{i,n}^\top R_n^{-1} \mathbf{X}_{i,n},$$

où  $R_n$  est l'estimation de la matrice de corrélation basée sur le tau de Kendall. À la limite quand  $n \rightarrow \infty$ , on s'attend à ce que  $\mathcal{G}_{n,n}^2, \dots, \mathcal{G}_{n,n}^2$  se comporte comme un échantillon i.i.d. de même loi que  $\mathcal{G}^2$  sous le modèle  $\mathcal{E}(0, R, g)$ .

### 4.3 Une statistique de Cramér–von Mises

La statistique d'adéquation proposée consiste à comparer la loi empirique de  $\mathcal{G}_{n,n}^2, \dots, \mathcal{G}_{n,n}^2$  à la loi  $\Psi_{\mathcal{G}^2}$  de  $\mathcal{G}^2$  qui est attendue sous  $\mathcal{H}_0$ . Soit donc

$$\Psi_{n,\mathcal{G}^2}(\gamma) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathcal{G}_{i,n}^2 \leq \gamma).$$

La distance de Cramér–von Mises entre  $\Psi_{n,\mathcal{G}^2}$  et  $\Psi_{\mathcal{G}^2}$  est définie par

$$V_n = n \int_0^\infty \{\Psi_{n,\mathcal{G}^2}(\gamma) - \Psi_{\mathcal{G}^2}(\gamma)\}^2 d\Psi_{\mathcal{G}^2}(\gamma).$$

Dans l'article présenté au Chapitre 5, on mentionne que

$$V_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{1 - \Psi_{R^2}(\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2)\} - \sum_{i=1}^n \{1 - \Psi_{R^2}(\mathcal{G}_{i,n}^2)\} + \frac{n}{3},$$

où  $a \vee b = \max(a, b)$ . La démonstration n'est toutefois pas fournie. En fait, en développant le carré, on a  $V_n = n(A_1 - 2A_2 + A_3)$ , où

$$\begin{aligned} A_1 &= \int_0^\infty \Psi_{n,\mathcal{G}^2}^2(\gamma) d\Psi_{\mathcal{G}^2}(\gamma), \\ A_2 &= \int_0^\infty \Psi_{n,\mathcal{G}^2}(\gamma) \Psi_{\mathcal{G}^2}(\gamma) d\Psi_{\mathcal{G}^2}(\gamma), \\ A_3 &= \int_0^\infty \Psi_{\mathcal{G}^2}^2(\gamma) d\Psi_{\mathcal{G}^2}(\gamma). \end{aligned}$$

Premièrement,

$$\begin{aligned}
A_1 &= \int_0^\infty \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathcal{G}_{i,n}^2 \leq \gamma) \right\}^2 d\Psi_{\mathcal{G}^2}(\gamma) \\
&= \int_0^\infty \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}(\mathcal{G}_{i,n}^2 \leq \gamma) \mathbf{1}(\mathcal{G}_{j,n}^2 \leq \gamma) d\Psi_{\mathcal{G}^2}(\gamma) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty \mathbf{1}(\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2 \leq \gamma) d\Psi_{\mathcal{G}^2}(\gamma) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2}^\infty d\Psi_{\mathcal{G}^2}(\gamma) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Psi_{\mathcal{G}^2}(\gamma) \Big|_{\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2}^\infty \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \{1 - \Psi_{\mathcal{G}^2}(\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2)\}.
\end{aligned}$$

Ensuite,

$$\begin{aligned}
A_2 &= \int_0^\infty \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathcal{G}_{i,n}^2 \leq \gamma) \Psi_{\mathcal{G}^2}(\gamma) d\Psi_{\mathcal{G}^2}(\gamma) \\
&= \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{G}_{i,n}^2}^\infty \Psi_{\mathcal{G}^2}(\gamma) d\Psi_{\mathcal{G}^2}(\gamma) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{\{\Psi_{\mathcal{G}^2}(\gamma)\}^2}{2} \Big|_{\mathcal{G}_{i,n}^2}^\infty \\
&= \frac{1}{2n} \sum_{i=1}^n \{1 - \Psi_{\mathcal{G}^2}(\mathcal{G}_{i,n}^2)\}.
\end{aligned}$$

Enfin,

$$A_3 = \int_0^\infty \Psi_{\mathcal{G}^2}^2(\gamma) d\Psi_{\mathcal{G}^2}(\gamma) = \frac{\{\Psi_{\mathcal{G}^2}(\gamma)\}^3}{3} \Big|_{-\infty}^\infty = \frac{1}{3}.$$

En rassemblant le tout via  $V_n = n(A_1 - 2A_2 + A_3)$ , on trouve la formule explicite annoncée.



## 4.4 Approximation des $p$ -valeurs

L'hypothèse nulle d'une distribution méta-elliptique  $\mathcal{ME}(R, g)$  sera acceptée pour de *petites* valeurs de la statistique  $V_n$ . Pour établir un critère de décision formel, il faudrait toutefois se baser sur la loi (asymptotique ou non) de  $V_n$ . D'emblée, on mentionne que cette loi a certainement une forme extrêmement compliquée. De plus, celle-ci dépend de la valeur inconnue de la matrice de corrélation. On doit donc envisager une méthode alternative pour l'obtention de valeurs critiques, ou de façon équivalente, de  $p$ -valeurs.

La stratégie employée sera basée sur la technique du *bootstrap* paramétrique. L'idée est de générer  $M$  échantillons indépendants (par exemple,  $M = 10\,000$ ) à partir de la loi elliptique sous  $\mathcal{H}_0$ . Ensuite, on calcule la statistique de test  $V_n$  pour chacun de ces échantillons *artificiel*. On se retrouve alors avec des *copies*  $V_n^{(1)}, \dots, V_n^{(M)}$  de la statistique de test, à partir desquelles une  $p$ -valeur est donnée par

$$\hat{p} = \frac{1}{M} \sum_{h=1}^M \mathbf{1}(V_n^{(h)} > V_n).$$

On rejette  $\mathcal{H}_0$  si  $\hat{p}$  est inférieure à un seuil fixé  $\alpha$ .

## CHAPITRE 5

# STATISTICAL PROCEDURES FOR THE SELECTION OF A MULTIDIMENSIONAL META-ELLIPTICAL DISTRIBUTION

Rachelle Bellerive & Jean-François Quessy

*Département de mathématiques et d'informatique, Université du Québec à  
Trois-Rivières, Trois-Rivières (QC) Canada, G9A 5H7*

**Abstract :** Meta-elliptical distributions are multivariate statistical models in which the dependence structure is governed by an elliptical copula and where the marginal distributions are arbitrary. In this paper, goodness-of-fit tests are proposed for the construction of an appropriate meta-elliptical model for multidimensional data. While the choice of the marginal distributions can be guided by classical goodness-of-fit testing, how to select an adequate elliptical copula is less clear. In order to fill this gap, formal copula goodness-of-fit methodologies are developed here around the so-called *radial part* that characterizes an elliptical distribution. The key idea consists in estimating its univariate distribution function from a pseudo-sample derived from the original multivariate observations. Then, a Cramér–von Mises distance between this non-parametric estimator and the expected parametric version under the null hypothesis is used as a test statistic. An approximate  $p$ -value is

obtained from an application of the parametric bootstrap. The method is extended to the case where the elliptical generator has unknown parameters using a minimum-distance criterion. The efficiency of the tests and related parameter estimation procedures is carefully investigated through Monte-Carlo simulations. The methods are illustrated on the Danish fire insurance, Upper Mississippi river, Oil currency and Uranium exploration data sets.

**Keywords :** Copula, goodness-of-fit test, meta-elliptical distributions, minimum-distance method, parametric bootstrap

## 5.1 Introduction

Following Cambanis et al. (1981), a random vector  $\mathbf{X} \in \mathbb{R}^d$  is said to be elliptically contoured if its associated characteristic function is of the form

$$\mathbb{E} \left( e^{i \mathbf{t}^\top \mathbf{X}} \right) = \exp \left( i \mathbf{t}^\top \boldsymbol{\mu} \right) \Upsilon \left( \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} \right), \quad (5.1)$$

where  $\mathbf{t} = (t_1, \dots, t_d)^\top$ ,  $\boldsymbol{\mu} \in \mathbb{R}^d$  is a mean vector and  $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$  is a positive definite matrix; the function  $\Upsilon : \mathbb{R}^+ \rightarrow \mathbb{R}$  is called the *characteristic generator* and is such that  $v(t_1, \dots, t_d) = \Upsilon(t_1^2 + \dots + t_d^2)$  is a  $d$ -variate characteristic function. This class of multidimensional distributions was first introduced by Kelker (1970) as a generalization of the classical Normal law, and comprises in particular the spherical laws when  $\boldsymbol{\mu} = (0, \dots, 0)^\top \in \mathbb{R}^d$  and  $\boldsymbol{\Sigma}$  is the identity matrix. The most notorious members of the general elliptical family are the multivariate Normal and Student distributions. These models have proven useful for statistical modeling in situations where alternatives to normality

were needed, especially in finance and hydrology. Their success lies mainly in that they allow for heterogeneous levels of dependence for the pairs, via the elements of a covariance matrix, and several kinds of tail behaviors by means of the elliptical generator. Among the many applications of elliptical distributions, one can cite Landsman & Valdez (2003) for the computation of tail conditional expectations, and Owen & Rabinovitch (1983) in the theory of portfolio choice.

Even if elliptical models are quite flexible, they nevertheless have the disadvantage that all marginal distributions have the same analytical form up to location and scale factors. This can be quite restrictive in situations where, for example, marginal tail behaviors are of a different nature, some components being heavy tailed and others having light tails. A model-building strategy that allows much more flexibility is to consider the *copula* of elliptically contoured distributions combined with any choice of the marginal distributions. Of course, the starting point of this approach is Sklar's Theorem, who ensures that if the marginal distributions of a  $d$ -variate random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top$  are continuous, then there exists a unique copula  $C_{\mathbf{X}} : [0, 1]^d \rightarrow [0, 1]$  such that for each  $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ ,

$$P(\mathbf{X} \leq \mathbf{x}) = C_{\mathbf{X}} \{P(X_1 \leq x_1), \dots, P(X_d \leq x_d)\}. \quad (5.2)$$

When  $\mathbf{X}$  is elliptically contoured, the function  $C_{\mathbf{X}}$  is called an *elliptical copula*. These dependence functions are the key elements of the multivariate meta-elliptical distributions, which are the multivariate probability laws whose underlying copula is elliptical. In other words, a random vector  $\mathbf{Y} = (Y_1, \dots, Y_d)^\top \in \mathbb{R}^d$  is said to be meta-elliptically distributed with marginal

distributions  $F_1, \dots, F_d$  if its copula  $C_Y$  belongs to the family of elliptical copulas. These models thus provide a general framework where one can select arbitrary margins and where the dependence structure is governed by a correlation matrix and an elliptical generator. As enlightened by Genest et al. (2007), “Meta-elliptical models are a good compromise between convenience and flexibility.” They appear in several contexts of multivariate statistical analysis. For example, Abdous et al. (2005) studied their dependence properties in the bivariate case, Landsman (2009) used them to model capital allocation and Wang et al. (2010) in multivariate modeling of hydrological data.

Despite the numerous successes in the application of meta-elliptical models, the problem of selecting an appropriate elliptical copula is still an open problem. Up until now, only the methodology based on the empirical copula process (see Genest et al., 2009) and on Kendall’s process (see Genest et al., 2006) have been suggested for the selection of a meta-elliptical model. This has been done by Genest et al. (2007), where trivariate hydrological data were modeled. However, the power properties of these tests have not been investigated for elliptical copulas. Furthermore, these methods have been considered solely in cases where the elliptical generator is entirely known, *i.e.* no parameter needs to be estimated.

In this paper, statistical tools for the construction of an appropriate meta-elliptical model are proposed and their properties in small samples are investigated. While the selection of suitable marginal distributions can be made by standard univariate goodness-of-fit methods (see Durbin, 1973, for example),

how to thoroughly choose an elliptical copula isn't as clear. In order to fill this gap, a special focus is put here on the investigation of new goodness-of-fit tests specifically designed to assess the quality of the fit of a given elliptical copula family on multivariate data. The key idea exploits a characterization of elliptically contoured random vectors via their so-called *radial part*. Both the cases where the radial part has an entirely known distribution and a distribution with unknown parameters are explored. The latter situation enables to test, for example, for a Student dependence structure without having to specify the number of degrees of freedom. A graphical tool based on sample counterparts of the radial part is also proposed.

This paper is organized as follows. In Section 2, standard results on elliptical and meta-elliptical families of distributions are reviewed; many members, including the Normal, Student and Pearson type II distributions are described. Section 3 concerns the development of a goodness-of-fit procedure for selecting an appropriate elliptical copula; simulation results indicate that the method works well. In Section 4, the framework is extended in order to consider cases where the elliptical generator has unknown parameters; the behavior of the proposed minimum-distance test statistic in small samples is numerically studied as well. Finally, Section 5 reports the results of the statistical analysis of four data sets in the light of the new tests.

## 5.2 The meta-elliptical family of dependence functions

### 5.2.1 Elliptically contoured distributions

Let  $\mathbf{X} \in \mathbb{R}^d$  be elliptically contoured, *i.e.* its characteristic function has the form (5.1). Following the same line of work as Schoenberg (1938) on spherically symmetric distributions, Cambanis et al. (1981) obtained that  $\mathbf{X}$  admits the stochastic representation

$$\mathbf{X} \stackrel{d}{=} \mu + \mathcal{G}A\mathcal{U}, \quad (5.3)$$

where  $\mu \in \mathbb{R}^d$  is the mean vector, the *radial part*  $\mathcal{G}$  is a positive random variable,  $A \in \mathbb{R}^{d \times d}$  is a (fixed) matrix such that  $AA^\top = \Sigma$  and  $\mathcal{U}$  is a random vector uniformly distributed on the unit sphere  $\mathbb{S}^{d-1}$  in  $\mathbb{R}^d$ , *i.e.*

$$\mathbb{S}^{d-1} = \{\mathbf{u} = (u_1, \dots, u_d)^\top \in \mathbb{R}^d : \mathbf{u}^\top \mathbf{u} = 1\}.$$

Note that the distribution of  $\mathcal{G}$  can be standardized in such a way that  $\Sigma$  corresponds to the covariance matrix, *i.e.*  $E(\mathbf{X}\mathbf{X}^\top) = \Sigma$ . From equation (5.3),

$$\|A^{-1}(\mathbf{X} - \mu)\|^2 = \mathcal{G}^2 \|\mathcal{U}\|^2 = \mathcal{G}^2, \quad (5.4)$$

where  $\|\cdot\|$  is the usual euclidian norm in  $\mathbb{R}^d$ . Since in practice, one generally observes *copies* of the random vector  $\mathbf{X}$ , representation (5.4) enables to *recover* the radial part  $\mathcal{G}$  that characterizes an elliptical distribution. This simple observation is at the base of the statistical methodologies developed in this paper. When  $\mathcal{G}$  has a density, the density of  $\mathbf{X}$  exists and is of the form

$$h_{\mu, \Sigma, g}(x_1, \dots, x_d) = |\Sigma|^{-1/2} g \left\{ \frac{1}{2} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right\}, \quad (5.5)$$

in terms of a *density generator*  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  standardized in such a way that

$$\int_0^\infty \gamma^{\frac{d}{2}-1} g(\gamma) \, d\gamma = \frac{\Gamma(d/2)}{(2\pi)^{d/2}}.$$

In the sequel, one notes  $\mathbf{X} \sim \mathcal{E}(\mu, \Sigma, g)$  whenever  $\mathbf{X}$  admits the representation (5.3).

By construction, the marginal distributions of elliptically contoured random vectors are identical. The latter, which will be denoted  $Q_g$  in the sequel, belongs to the univariate elliptical family ; these symmetric models are particular cases of location-scale distributions. Using the transformation in polar coordinates described in the Appendix of the article by Landsman & Valdez (2003), one deduces that the density associated to  $Q_g$  is proportional to

$$q_g(x) = \int_{x^2}^\infty (y-x)^{\frac{d-3}{2}} g\left(\frac{y}{2}\right) dy.$$

It is clear from (5.5) that elliptical distributions are symmetric about  $\mu$ . This formula also entails that the density of  $\Lambda^{-1}(\mathbf{X} - \mu)$  is  $g(\mathbf{x}^\top \mathbf{x}/2)$ , from which it follows that the density and distribution functions of the squared radial part  $\mathcal{G}^2$  are respectively

$$\psi_{\mathcal{G}^2}(\gamma) = \frac{\pi^{d/2}}{\Gamma(d/2)} \gamma^{\frac{d}{2}-1} g\left(\frac{\gamma}{2}\right) \quad \text{and} \quad \Psi_{\mathcal{G}^2}(\gamma) = \frac{\pi^{d/2}}{\Gamma(d/2)} \int_0^\gamma s^{\frac{d}{2}-1} g\left(\frac{s}{2}\right) ds.$$

Some of the most popular elliptical distributions are now described ; see Fang et al. (1990) for more details.

**Example 5.1** (Normal and Student distributions). *The multivariate Normal distribution, noted  $N$ , arises when*

$$g(\gamma) = \frac{1}{(2\pi)^{d/2}} \exp(-\gamma).$$



In that case,  $\mathcal{G}^2$  is chi-squared distributed with  $d$  degrees of freedom and the marginal distribution  $Q_g$  in the standard case corresponds to that of the  $\mathcal{N}(0, 1)$  law, namely

$$Q_g(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds.$$

A generalization of the Normal distribution often used in robustness studies of Gaussian-based statistical methods (see Bilodeau & Brenner, 1999, for instance) is the multivariate Student distribution. In that case, the elliptical generator is proportional to

$$g_\nu(\gamma) = \left(1 + \frac{\gamma}{\nu}\right)^{-\frac{\nu}{2}-1},$$

where  $\nu > 0$  is often referred to as the number of degrees of freedom. The law of  $d\mathcal{G}^2$  is then the Fisher–Snedecor distribution with  $d$  and  $\nu$  degrees of freedom. The Student distribution will be noted  $T_\nu$  for the remaining of the paper. The case  $\nu = 1$  corresponds to the Cauchy distribution.

**Example 5.2** (Normal variance mixture models). The  $N$  and  $T_\nu$  distributions can be viewed as particular cases of the general normal variance mixture models described, for example, by Klüppelberg & Kuhn (2009). Random variables in this class admit the stochastic representation  $\mathbf{X} = \sqrt{W} \mathbf{A} \mathbf{Z}$ , where  $\mathbf{Z}$  follows a  $d$ -dimensional standard Normal distribution and  $W$  is a non-negative random variable. One can then show that  $\mathcal{G}^2 = W \mathbf{Z}^\top \mathbf{Z} / \mathbb{E}(W)$ , so that  $\mathcal{G}^2 \stackrel{d}{=} W \chi_d^2 / \mathbb{E}(W)$ . The Normal and Student laws appear when  $W \equiv 1$  and  $1/W \stackrel{d}{=} \chi_\nu^2 / \nu$ , respectively. Other popular models also possess this stochastic representation. For example, one recovers the multivariate Laplace distribution when  $W$  is an exponential random variable; for more details, see e.g. Eltoft et al. (2006).

**Exemple 5.3** (Pearson type II). *An elliptical distribution is said to be unbounded if any subset of  $\mathbb{R}^d$  has non-null probability; this is the case, in particular, for the Normal and Student distributions. The bounded case happens whenever the elliptical generator is such that  $g(\gamma) = 0$  for all  $\gamma$  exceeding some constant  $K > 0$ . This property is shared by the Pearson type II distribution whose generator is*

$$g_\theta(\gamma) = \frac{\Gamma\left(\frac{d}{2} + \theta + 1\right)}{\pi^{d/2} \Gamma(\theta + 1)} (1 - 2\gamma)^\theta, \quad \gamma \in [0, 1/2],$$

where  $\theta > -1$ . One can show that the associated distribution of  $\mathcal{G}^2$  is Beta with parameters  $d/2$  and  $\theta + 1$ . This model will be referred to as  $Pe_\theta$  in the sequel.

**Exemple 5.4** (Exponential power family). *Consider an elliptical generator proportional to*

$$g_{\theta_1, \theta_2}(\gamma) = \exp(-\theta_1 \gamma^{\theta_2}),$$

where  $\theta_1, \theta_2 > 0$ . One recovers the Normal distribution when  $\theta_1 = \theta_2 = 1$  and Laplace's distribution when  $\theta_1 = \sqrt{2}$ ,  $\theta_2 = 1/2$ . Another interesting sub-model is Kotz's distribution (see Kotz, 1975) that arises when  $\theta_1 = \theta$  and  $\theta_2 = 1$ , i.e.  $g_\theta(\gamma) \propto \exp(-\theta \gamma)$ . It is worth noting that  $\theta_1$  only acts here as a scale parameter, so that it has no influence on the underlying copula. The copula associated with  $g_{\theta_1, \theta_2}$  is then the same as the one generated by

$$g_\theta(\gamma) \propto \exp(-\gamma^\theta).$$

In that case,  $\mathcal{G}^2 \stackrel{d}{=} Y^{1/\theta}$ , where  $Y$  is Gamma distributed with parameters  $d/(2\theta)$  and  $2^\theta$ .

### 5.2.2 Elliptical copulas and meta-elliptical models

Considering the *inverse* version of Sklar's representation in (5.2), one deduces that the unique copula associated to an elliptically contoured random vector  $\mathbf{X} = (X_1, \dots, X_d)^\top \sim \mathcal{E}(\mu, \Sigma, g)$  is

$$C_{\mathbf{X}}(\mathbf{u}) = \mathbb{P} \{ \mathbf{X} \leq \mathbf{Q}_g^{-1}(\mathbf{u}) \}, \quad \mathbf{u} = (u_1, \dots, u_d)^\top,$$

where

$$\mathbf{Q}_g^{-1}(\mathbf{u}) = (Q_g^{-1}(u_1), \dots, Q_g^{-1}(u_d))^\top$$

is the vector of componentwise inverses of the marginal distributions. Since copulas are invariant under strictly increasing transformations of the individual variables, the copula of  $\mathbf{X} \stackrel{d}{=} \mu + \mathcal{GAU}$  is the same as the copula of  $\tilde{\mathbf{X}} \stackrel{d}{=} \mathcal{GAU}$ , where  $\tilde{A}\tilde{A}^\top = R$  is a correlation matrix. Hence, without loss of generality, one can assume that  $\mathbf{X} \sim \mathcal{E}(\mathbf{0}, R, g)$  in the sequel. From (5.5), one then deduces that the copula of  $\mathbf{X}$  is

$$C_{R,g}(\mathbf{u}) = |R|^{-1/2} \int_{-\infty}^{Q_g^{-1}(u_1)} \cdots \int_{-\infty}^{Q_g^{-1}(u_d)} g\left(\frac{1}{2} \mathbf{x}^\top R^{-1} \mathbf{x}\right) d\mathbf{x}, \quad (5.6)$$

from which it follows that the density of an elliptical copula is of the form

$$c_{R,g}(\mathbf{u}) = |R|^{-1/2} g\left(\frac{1}{2} \tilde{\mathbf{x}}^\top R^{-1} \tilde{\mathbf{x}}\right),$$

where  $\tilde{\mathbf{x}} = \mathbf{Q}_g^{-1}(\mathbf{u})$ . In Figure 5.1, one can see the copula density plots (*i.e.* the normalized ranks) from the realization of 5 000 pairs from the  $T_1, T_3, T_9, N, \text{Pe}_1$  and  $\text{Pe}_5$  elliptical distributions.

FIGURE 5.1 ABOUT HERE

Elliptical copulas lead directly to the meta-elliptical family of models. Indeed, a random vector  $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$  is said to have a meta-elliptical distribution with correlation matrix  $R$ , generator  $g$  and marginals  $\mathbf{F} = (F_1, \dots, F_d)^\top$  if its underlying copula is  $C_{R,g}$ ; it will be noted  $\mathbf{Y} \sim \mathcal{ME}(R, g, \mathbf{F})$  in the sequel. The meta-Gaussian distribution have been first considered by Kelly & Krzysztofowicz (1997), and the idea was extended to general meta-elliptical distributions by Fang et al. (2002).

The multidimensional probability integral transformation ensures that  $\mathbf{F}(\mathbf{Y}) \sim C_{R,g}$ , from which it follows that

$$\tilde{\mathbf{X}} = \mathbf{Q}_g^{-1} \circ \mathbf{F}(\mathbf{Y}) \sim \mathcal{E}(\mathbf{0}, R, g). \quad (5.7)$$

Therefore, a particular case of equation (5.4) with  $\mu = (0, \dots, 0)^\top \in \mathbb{R}^d$  yields

$$\tilde{\mathcal{G}}^2 = \tilde{\mathbf{X}}^\top R^{-1} \tilde{\mathbf{X}} \sim \Psi_{\mathcal{G}^2}. \quad (5.8)$$

Equation (5.8) is at the basis of the goodness-of-fit procedures developed in the next two sections.

### 5.3 Goodness-of-fit procedure in the case of a fixed generator

#### 5.3.1 Context

Let  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , where  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})^\top$ , be independent copies of a random vector  $\mathbf{Y} \in \mathbb{R}^d$ . The aim of this section is to develop a statistical method for the null and alternative hypotheses

$$\mathcal{H}_0 : \mathbf{Y} \sim \mathcal{ME}(R, g, \mathbf{F}) \quad \text{and} \quad \mathcal{H}_1 : \mathbf{Y} \approx \mathcal{ME}(R, g, \mathbf{F}), \quad (5.9)$$

where  $R$  and  $\mathbf{F}$  are unknown, while the elliptical generator  $g$  is entirely known. This setting encompasses the cases where one wants to test for a meta-Gaussian distribution or a meta-Student distribution with a fixed number of degrees of freedom. In Section 5.4, the assumption of a fixed generator is relaxed in order to allow for parametric generators, *i.e.*  $g = g_\theta$  for some  $\theta \in \Theta$ .

The fact that the vector  $\mathbf{F}$  is undetermined enables to infer on the dependence structure of  $\mathbf{Y}$  regardless of the marginal behavior of its components. In general, non-parametric estimation of the marginal distributions uses empirical distribution functions, *i.e.*  $\mathbf{F}$  is estimated by  $\mathbf{F}_n = (F_{n1}, \dots, F_{nd})^\top$ , where for each  $j \in \{1, \dots, d\}$ ,

$$F_{nj}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_{ij} \leq y).$$

A model-free estimation of the correlation matrix  $R$  can also be accomplished. This is described in the next subsection.

### 5.3.2 Estimation of the correlation matrix

An interesting feature of elliptical copulas is the fact that the correlation matrix can be estimated independently of the form of the elliptical generator  $g$ . The idea is to exploit a relationship that exists between the entries of  $R$  and the value of Kendall's measure of association for each of the  $d(d-1)/2$  pairs. Recall that Kendall's tau for a random pair  $(Y_1, Y_2)$  is defined by

$$\tau(Y_1, Y_2) = P\{(Y_{11} - Y_{21})(Y_{12} - Y_{22}) > 0\} - P\{(Y_{11} - Y_{21})(Y_{12} - Y_{22}) < 0\},$$

where  $(Y_{11}, Y_{12})$  and  $(Y_{21}, Y_{22})$  are independent copies of  $(Y_1, Y_2)$ . For a meta-elliptically contoured random vector  $\mathbf{Y} \sim \mathcal{ME}(R, g, \mathbf{F})$ , one deduces from Lindskog et al. (2003) and Fang et al. (2002) that

$$\tau(Y_k, Y_\ell) = \frac{2}{\pi} \sin^{-1}(R_{k\ell})$$

for each  $(k, \ell)$  such that  $k < \ell \in \{1, \dots, d\}$ . Then, if  $\tau_{n,k\ell}$  is an empirical version of  $\tau(Y_k, Y_\ell)$ , the correlation matrix  $R$  can easily be estimated by  $R_n = (R_n)_{k\ell}$ , where

$$(R_n)_{k\ell} = \sin\left(\frac{\pi}{2} \tau_{n,k\ell}\right). \quad (5.10)$$

Usually,  $\tau_{n,k\ell}$  is the U-statistic defined by

$$\tau_{n,k\ell} = -1 + \frac{4}{n(n-1)} \sum_{i < j} \{\mathbf{1}(Y_{ki} < Y_{kj}, Y_{\ell i} < Y_{\ell j}) + \mathbf{1}(Y_{ki} > Y_{kj}, Y_{\ell i} > Y_{\ell j})\},$$

which is unbiased for  $\tau(Y_k, Y_\ell)$  and asymptotically Normal as  $n \rightarrow \infty$ ; see the monograph by Lee (1990) for a complete survey of the theory of U-statistics. The estimation method in (5.10) was employed by Genest et al. (2007) and

Klüppelberg & Kuhn (2009). As a generalization of the large-sample behavior of Kendall's tau for a single pair, Klüppelberg & Kuhn (2009) obtained that

$$\sqrt{n}(R_n - R) \rightsquigarrow \mathcal{N}_{\frac{d(d-1)}{2}}(\mathbf{0}, \Lambda)$$

for some asymptotic variance-covariance matrix  $\Lambda$  that depends on the parameters  $R$  and  $g$  of the underlying elliptical copula  $C_{R,g}$ . In Table 5.1, the results of a simulation study that aims to evaluate the efficiency of  $R_n$  as an estimator of  $R$  are presented. To this end, let  $S \in \mathbb{R}^{d \times d}$  be a symmetric matrix and define the mean of its off-diagonal elements by

$$\mathcal{L}(S) = \frac{2}{d(d-1)} \sum_{k < \ell \in \{1, \dots, d\}} S_{k\ell}.$$

The mean-squared error criterion that was used in the investigation is

$$\text{MSE}_R(R_n) = \text{E} \{ (\mathcal{L}(R_n) - \mathcal{L}(R))^2 \}$$

and has been estimated by

$$\widehat{\text{MSE}}_R(R_n) = \frac{1}{N} \sum_{i=1}^N \{ \mathcal{L}(R_{ni}) - \mathcal{L}(R) \}^2,$$

where for each  $i \in \{1, \dots, N\}$ , the sample correlation matrix  $R_{ni}$  is computed from a random sample of size  $n$  from the  $\mathcal{E}(\mathbf{0}, R, g)$  distribution for a given elliptical model generated by  $g$ . In our empirical study, only the equi-correlated case has been considered, *i.e.*  $R_{k\ell} = \rho$  for all  $k < \ell \in \{1, \dots, d\}$ . Note that the entries in Table 5.1 correspond to the estimated *standardized* MSE, *i.e.*  $n \times \widehat{\text{MSE}}_R(R_n)$ . One can see that the standardized MSE's

- (i) are quite equivalent for  $n = 100$  and  $n = 250$ ;

- (ii) decrease monotonically as the strength of the pairwise dependence coefficient  $\rho$  increases when  $d \in \{2, 3\}$ ; for  $d \in \{4, 5\}$ , they are equivalent when  $\rho \in \{1/4, 1/2\}$  and about half lower when  $\rho = 3/4$ ;
- (iii) tend to be lower as the dimension  $d$  increases;
- (iv) depend on the elliptical generator : the lowest values appear for the Pearson type II distribution, while under the Student model, they are higher for low values of  $\nu$  and tend to the values for the Normal distribution when  $\nu \in \{6, 9\}$ , as was expected.

TABLE 5.1 ABOUT HERE

### 5.3.3 Test statistic and parametric bootstrap

Under the null hypothesis  $\mathcal{H}_0$  stated in equation (5.9), each component of the random sample  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  follows a  $\mathcal{ME}(R, g, \mathbf{F})$  distribution. Hence, from equation (5.7), one has for each  $i \in \{1, \dots, n\}$  that

$$\tilde{\mathbf{X}}_i = \mathbf{Q}_g^{-1} \circ \mathbf{F}(\mathbf{Y}_i) \sim \mathcal{E}(\mathbf{0}, R, g).$$

Since the vector of the marginal distributions  $\mathbf{F}$  is unknown,  $\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_n$  are unobservable. One would rather work with the *pseudo-sample*  $\mathbf{X}_{1,n}, \dots, \mathbf{X}_{n,n}$ , where for each  $i \in \{1, \dots, n\}$ ,

$$\mathbf{X}_{i,n} = \mathbf{Q}_g^{-1} \circ \mathbf{F}_n(\mathbf{Y}_i). \quad (5.11)$$

Note that

$$\mathbf{F}_n(\mathbf{Y}_i) = \left( \frac{\text{Rank}(Y_{i1})}{n}, \dots, \frac{\text{Rank}(Y_{id})}{n} \right)^\top,$$



where  $\text{Rank}(Y_{ij})$  is the rank of  $Y_{ij}$  among  $Y_{1j}, \dots, Y_{nj}$ ; it is then clear that the upcoming statistical methods are entirely rank-based. Since  $\mathbf{F}_n$  is a uniformly consistent estimator of  $\mathbf{F}$ , it is expected that  $\mathbf{X}_{i,n}$  follows *approximately* a  $\mathcal{E}(\mathbf{0}, R, g)$  distribution. The construction of sample counterparts of the random variable  $\mathcal{G}^2$  adds another level of complexity since  $R$  must be estimated, too. Starting from equation (5.8) and admitting that

$$\mathcal{G}_{i,n}^2 = \mathbf{X}_{i,n}^\top R_n^{-1} \mathbf{X}_{i,n} \quad (5.12)$$

for each  $i \in \{1, \dots, n\}$ , one then has a pseudo-sample  $\mathcal{G}_{1,n}^2, \dots, \mathcal{G}_{n,n}^2$  that should behave asymptotically like the random variable  $\mathcal{G}^2$ . Showing thoroughly that it is indeed the case is highly non-trivial and possibly requires asymptotic tools developed by Ghoudi & Rémillard (2004) for multivariate pseudo-observations. A valid non-parametric sample version of  $\Psi_{\mathcal{G}^2}$  would then be given by

$$\Psi_n(\gamma) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathcal{G}_{i,n}^2 \leq \gamma).$$

Reformulating the null and alternative hypotheses as  $\mathcal{H}_0 : \mathcal{G}^2 \sim \Psi_{\mathcal{G}^2}$  and  $\mathcal{H}_1 : \mathcal{G}^2 \not\sim \Psi_{\mathcal{G}^2}$ , where  $\Psi_{\mathcal{G}^2}$  is associated to the  $\mathcal{E}(\mathbf{0}, R, g)$  distribution, a naturel test consist in rejecting  $\mathcal{H}_0$  for large values of the Cramér–von Mises distance between  $\Psi_n$  and  $\Psi_{\mathcal{G}^2}$ , namely

$$V_n = n \int_0^\infty \{\Psi_n(\gamma) - \Psi_{\mathcal{G}^2}(\gamma)\}^2 \psi_{\mathcal{G}^2}(\gamma) d\gamma.$$

One can easily show that

$$V_n = \frac{n}{3} + \sum_{i=1}^n \Psi_{\mathcal{G}^2}(\mathcal{G}_{i,n}^2) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{\Psi_{\mathcal{G}^2}(\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2)\}^2.$$

A test statistic similar to  $V_n$  was proposed by Malevergne & Sornette (2003) in the special case of the Normal copula.

The  $p$ -value of the test based on  $V_n$  will be computed by an application of the parametric bootstrap method described next. Showing the asymptotic validity, as  $n \rightarrow \infty$ , of the algorithm below could probably be obtained from results by Genest & Rémillard (2008). Based on the simulation results presented in subsection 5.3.4, however, there is every reason to believe that the method works well.

**Algorithm 5.1.** *Given a random sample  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , the parametric bootstrap under the null hypothesis of a meta-elliptical distribution  $\mathcal{ME}(R, g, \mathbf{F})$  consists in*

- (1) *computing the sample correlation matrix  $R_n$  and the test statistic  $V_n$  from  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ ;*
- (2) *generating, for a sufficiently large  $M \in \mathbb{N}$ , independent random samples*

$$\left( \mathbf{Y}_1^{(1)}, \dots, \mathbf{Y}_n^{(1)} \right), \dots, \left( \mathbf{Y}_1^{(M)}, \dots, \mathbf{Y}_n^{(M)} \right)$$

*from the  $\mathcal{E}(0, R_n, g)$  distribution, and computing  $V_n^{(1)}, \dots, V_n^{(M)}$  associated to each of these samples;*

- (3) *obtaining the approximate  $p$ -value*

$$p_{n,M} = \frac{1}{M} \sum_{h=1}^M \mathbf{1} (V_n^{(h)} > V_n).$$

An approximate confidence interval for  $\Psi_{\mathcal{G}^2}$  can be built from Algorithm 5.1. Suppose one is seeking for a  $q_\alpha \in \mathbb{R}^+$  such that

$$\lim_{n \rightarrow \infty} \mathbb{P} (\|\Psi_n - \Psi_{\mathcal{G}^2}\|_\infty \leq q_\alpha) = 1 - \alpha,$$

where  $1 - \alpha$  is the desired confidence level. Take  $\widehat{q}_\alpha$  as the  $100 \times (1 - \alpha)$ -th percentile of

$$\|\Psi_n^{(1)} - \Psi_{\mathcal{G}^2}\|_\infty, \dots, \|\Psi_n^{(M)} - \Psi_{\mathcal{G}^2}\|_\infty,$$

where  $\Psi_n^{(h)}$  is the empirical distribution function of  $\mathcal{G}_{1,n}^{(h)}, \dots, \mathcal{G}_{n,n}^{(h)}$  computed from the parametric bootstrap sample  $\mathbf{Y}_1^{(h)}, \dots, \mathbf{Y}_n^{(h)}$ . Then, the confidence band is defined for each  $x \in \mathbb{R}^+$  by

$$\mathbb{CB}_\alpha(x) = [\Psi_n(x) - \widehat{q}_\alpha, \Psi_n(x) + \widehat{q}_\alpha].$$

**Remarque 5.1.** *It is possible to extend the statistical method in this section to cases when the form of  $\Psi_{\mathcal{G}^2}$  in the model  $\mathcal{E}(\mathbf{0}, R, g)$  is not explicit. The idea is to replace the test statistic  $V_n$  by*

$$V_{n,N} = n \int_0^\infty \{\Psi_n(x) - \Psi_N(x)\}^2 dx,$$

where  $\Psi_N$  is the empirical distribution function of  $\mathcal{G}_{1,N}, \dots, \mathcal{G}_{N,N}$ , where for each  $i \in \{1, \dots, N\}$ ,  $\mathcal{G}_{i,N} = \mathbf{Y}_i^\top \mathbf{Y}_i$  and  $\mathbf{Y}_i \sim \mathcal{E}(\mathbf{0}, I_d, g)$ , with  $I_d \in \mathbb{R}^{d \times d}$  the identity matrix. Note that one could also use the weight function  $d\Psi_N$  in the definition of  $V_{n,N}$ , but  $dx$  is chosen here for computational convenience. In that case, one obtains a simple formula, namely

$$\begin{aligned} V_{n,N} &= \frac{2}{N} \sum_{i=1}^n \sum_{j=1}^N (\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,N}^2) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (\mathcal{G}_{i,n}^2 \vee \mathcal{G}_{j,n}^2) \\ &\quad - \frac{1}{n} \sum_{i=1}^N \sum_{j=1}^N (\mathcal{G}_{i,N}^2 \vee \mathcal{G}_{j,N}^2). \end{aligned}$$

The  $p$ -value can then be computed from a slight modification of Algorithm 5.1.

### 5.3.4 Investigation of the size and power of the test

The size and power of the test based on  $V_n$  have been investigated with the help of Monte-Carlo simulations. The elliptical copulas considered under the null hypothesis are those extracted from the  $T_1$ ,  $T_3$ ,  $T_6$ ,  $T_9$ ,  $N$ ,  $\text{Pe}_1$ ,  $\text{Pe}_2$  and  $\text{Pe}_5$  distributions. The probability of rejecting the null hypothesis is always estimated from 1 000 replicates and  $M = 1\,000$  bootstrap samples.

One can find the results for the bivariate case ( $d = 2$ ) when  $n = 100$  and  $n = 250$  in Table 5.2 and Table 5.3 respectively. In addition to the eight elliptical models listed above, the non-elliptical Clayton (CL) and Gumbel–Hougaard (GH) copulas were used as alternatives ; see Nelsen (2006) for more details on these two models. For the elliptical distributions, the parameter  $\rho \in \{1/4, 1/2, 3/4\}$  corresponds to the correlation coefficient, *i.e.*  $R_{21} = R_{12} = \rho$ , while for the Clayton and Gumbel–Hougaard models,  $\rho$  corresponds to Spearman’s rank correlation coefficient. The latter depends on the copula  $C$  of a bivariate population via  $\rho_S = 12 \int_0^1 \int_0^1 \{C(u, v) - uv\} du dv$ .

TABLE 5.2 AND TABLE 5.3 ABOUT HERE

One first sees that the test keeps its 5% nominal level quite well under most of the considered scenarios. The power of the test is also generally very good, which shows that the method discriminates well between the various elliptical models. Of course, the power increases as  $n$  increases. As expected, similar models are harder to distinguish, for example  $T_3$  vs  $T_6$ ,  $T_9$  vs  $N$  and  $\text{Pe}_1$  vs  $\text{Pe}_2$ . It is to note that the value of  $\rho$  has little influence on the power

results. The alternative model which is the most easily rejected is  $T_1$ , even when the null hypothesis is the closely related  $T_3$  distribution. As expected, two Student distributions are harder to distinguish when their associated number of degrees of freedom are relatively large, since then they are both close to the Normal distribution. Finally, the probability of rejecting the null hypothesis when the data come from the Clayton or the Gumbel–Hougaard copulas are generally high when the null hypothesis is one of the Pearson type II distributions. For some reason, these departures from the null hypothesis are not well detected when  $\mathcal{H}_0$  is the  $T_6$  or  $T_9$  model.

The multivariate case has also been considered. In Table 5.4, the results are presented for  $d \in \{3, 4\}$  and the same alternatives that were used for the results presented in Tables 5.2–5.3, except that the Clayton and Gumbel–Hougaard copulas are now excluded. Only the case when  $n = 100$  and  $R$  is equi-correlated are presented here. Many of the comments stated for the results Tables 5.2–5.3 apply here as well : the test keep its 5% nominal level quite well, and the probabilities of rejection are similar for the values of  $\rho$  that were considered. For a given scenario under  $\mathcal{H}_1$ , the power tends to be higher in dimension  $d = 4$  compared to  $d = 3$ . Finally note that the  $T_1$  model is highly rejected for all the null hypotheses.

TABLE 5.4 ABOUT HERE

## 5.4 Goodness-of-fit procedure in the case of a parametric generator

### 5.4.1 A minimum-distance method

Suppose that the generator of an elliptical distribution depends on some unknown parameter  $\theta \in \Theta \subseteq \mathbb{R}^p$ , i.e.  $g = g_\theta$ . In that case, the null and alternative hypotheses are

$$\mathcal{H}_0^* : \mathbf{Y} \sim \mathcal{ME}(R, g_\theta, \mathbf{F}) \quad \text{and} \quad \mathcal{H}_1^* : \mathbf{Y} \sim \mathcal{ME}(R, g_\theta, \mathbf{F}),$$

where  $R$ ,  $\theta$  and  $\mathbf{F}$  are unknown. In what follows, one writes  $Q_\theta$  for the marginal distributions of the elliptical law generated by  $g_\theta$ . Defining  $\mathbf{Q}_\theta^{-1} = (Q_\theta^{-1}, \dots, Q_\theta^{-1})^\top$ , a *parametric* version of (5.11) is then given by

$$\mathbf{X}_{i,n}(\theta) = \mathbf{Q}_\theta^{-1} \circ \mathbf{F}_n(\mathbf{Y}_i),$$

yielding the pseudo-sample  $\mathbf{X}_{1,n}(\theta), \dots, \mathbf{X}_{n,n}(\theta)$ . Similarly, a parametric version of (5.12) is

$$\mathcal{G}_{i,n}^2(\theta) = \mathbf{X}_{i,n}(\theta)^\top R_n^{-1} \mathbf{X}_{i,n}(\theta),$$

where  $R_n$  is the estimator described in subsection 5.3.2. For a fixed value of  $\theta$ , an empirical version of the distribution  $\Psi_{\mathcal{G}^2}(\gamma, \theta)$  of the squared radial part  $\mathcal{G}^2$  under the model  $\mathcal{E}(\mathbf{0}, R, g_\theta)$  is

$$\Psi_n(\gamma, \theta) = \frac{1}{n} \sum_{i=1}^n \mathbf{1} \{ \mathcal{G}_{i,n}^2(\theta) \leq \gamma \}.$$

Since  $\theta \in \Theta$  is unknown, the proposed goodness-of-fit procedure will be based on the minimum-distance statistic

$$W_n = \inf_{\theta \in \Theta} W_n(\theta),$$

where

$$\begin{aligned} W_n(\theta) &= n \int_0^\infty \{\Psi_n(\gamma, \theta) - \Psi_{\mathcal{G}^2}(\gamma, \theta)\}^2 \psi_{\mathcal{G}^2}(\gamma, \theta) d\gamma \\ &= \frac{n}{3} + \sum_{i=1}^n \Psi_{\mathcal{G}^2} \{ \mathcal{G}_{i,n}^2(\theta), \theta \} - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n [\Psi_{\mathcal{G}^2} \{ \mathcal{G}_{i,n}^2(\theta) \vee \mathcal{G}_{j,n}^2(\theta) \}, \theta]^2 \end{aligned}$$

is the Cramér–von Mises distance between  $F_n(\gamma, \theta)$  and  $F_{\mathcal{G}^2}(\gamma, \theta)$ . The asymptotic behavior of such statistics has been investigated in details by Pollard (1980). A variant of this method in a copula context has been studied by Tsukahara (2005). In practice,  $W_n$  will be approximated on a grid  $(\theta_1, \dots, \theta_T)$  of  $\Theta$  in such a way that

$$W_n \approx \min_{\theta_1, \dots, \theta_T \in \Theta} W_n(\theta).$$

An estimator of  $\theta$  that is implicit in the definition of  $W_n$  is

$$\theta_n = \arg \min_{\theta \in \Theta} W_n(\theta). \quad (5.13)$$

The performance of  $\theta_n$  as an estimator of  $\theta$  is investigated in subsection 5.4.3. For the test based on  $W_n$ , an extension of the parametric bootstrap described in Algorithm 5.1 is needed to compute a  $p$ -value; it is described next.

**Algorithm 5.2.** *Given a random sample  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ , the parametric bootstrap under the null hypothesis of a meta-elliptical distribution  $\mathcal{ME}(R, g_\theta, \mathbf{F})$  consists in*

- (1) *computing the sample correlation matrix  $R_n$ , the estimator  $\theta_n$  and the test statistic  $W_n$  from  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ ;*

(2) *generating, for a sufficiently large  $M \in \mathbb{N}$ , independent random samples*

$$\left( \mathbf{Y}_1^{(1)}, \dots, \mathbf{Y}_n^{(1)} \right), \dots, \left( \mathbf{Y}_1^{(M)}, \dots, \mathbf{Y}_n^{(M)} \right)$$

*from the  $\mathcal{E}(0, R_n, g_{\theta_n})$  distribution, and computing the test statistics  $W_n^{(1)}, \dots, W_n^{(M)}$  associated to each of these samples;*

(3) *obtaining the approximate  $p$ -value*

$$p_{n,M}^* = \frac{1}{M} \sum_{h=1}^M \mathbf{1}(W_n^{(h)} > W_n) .$$

In the light of the empirical results presented in the next subsection, this algorithm works well.

#### 5.4.2 Investigation of the power of the test

The power of the test based on  $W_n$  has been investigated when  $\mathcal{H}_0^*$  is either the Student or the Pearson type II copula. The results for  $n \in \{50, 100\}$  and  $d \in \{2, 3\}$  are in Table 5.5. When  $d = 3$ , the alternative models are restricted to the equi-correlated case  $R_{12} = R_{13} = R_{23} = 1/2$ .

TABLE 5.5 ABOUT HERE

One can say that the tests keep their 5% nominal level very well when the null hypothesis is a Student copula. The tests are not as good under Pearson type II alternatives when  $n = 50$ , but it improves markedly when  $n = 100$ . A problem seems to occur under Normal alternatives, where all tests should



reject at a rate close to 5%, which is not the case. It can be explained by the fact that the test statistic  $W_n$  is evaluated on a grid, so the limiting Normal case which occurs for  $\theta \rightarrow \infty$  is never attained.

Concerning the power of the tests, it obviously increases as the sample size increases. It is very high when testing for a Pearson type II copula under Student alternatives, both in the bivariate and in the trivariate case. Of course, the power decreases as the number of degrees of freedom increases, since then it gets closer to the limiting Normal case. The power of the test is not as good when testing for a Student copula under Pearson type II alternatives. It is significantly higher in dimension  $d = 3$  compared to  $d = 2$ .

### 5.4.3 Performance of the minimum-distance estimator

The performance of the minimum-distance estimator  $\theta_n$  described in equation (5.13) has been investigated when the underlying models are the one-parameter Student and Pearson type II distributions. To this end,  $N = 1\,000$  random samples of size  $n = 50$  have been simulated under the  $T_1, T_3, T_6, T_9, \text{Pe}_1, \text{Pe}_2$  and  $\text{Pe}_5$  copulas in dimensions  $d \in \{2, 3, 4\}$ . The entries in Table 5.6 are the estimated standardized mean-squared errors  $n \times \widehat{\text{MSE}}_\theta(\theta_n)$ , where

$$\widehat{\text{MSE}}_\theta(\theta_n) = \frac{1}{N} \sum_{i=1}^N (\theta_{ni} - \theta)^2$$

and  $\theta_{n1}, \dots, \theta_{nN}$  are computed from the simulated random samples. While it would be more interesting to compare these standardized mean-squared errors with those of competing estimators, some comments can nevertheless be made. First, one sees that they are smaller when  $d$  is large. Also, under the

Student copula, they increase markedly as the number of degrees of freedom increases; the inverse phenomenon occurs under Pearson type II alternatives, where the standardized mean-squared errors decrease as  $\theta$  increases.

TABLE 5.6 ABOUT HERE

## 5.5 Illustrations on real data

### 5.5.1 The Danish fire insurance data

These data consists of 2 167 insurance claims relative to fire losses collected at the Copenhagen Reinsurance Company covering the years 1980–1990. This data set has been considered by Rytgaard (1996) and Embrechts et al. (1997), among others. Here, an observation refers to a loss of property and a loss of contents. When only the non-zero values are taken into account, this results in a bivariate data set of  $n = 604$  pairs. The scatter plot and density plot of the normalized ranks are presented on the top panels of Figure 5.2.

FIGURE 5.2 ABOUT HERE

In order to find a suitable elliptical copula for these data, the test based on the statistic  $V_n$  has been applied for various models; the results of the analysis are presented in the left part of Table 5.8. From the entries therein, the  $T_6$ ,  $T_9$  et Normal models are not rejected at the 5% significance level. These results are somewhat confirmed by the test based on  $W_n$ , where the

Student model is highly accepted ( $W_n = 0.0208$ ,  $p_{n,M}^* = 0.979$ ,  $\theta_n = 10.75$ ), while the Pearson type II copula is also accepted, but with a much smaller  $p$ -value ( $p_{n,M}^* = 0.088$ ). In the latter case, the fact that  $\theta_n = 15$  indicates that a Normal dependence structure would probably be suitable.

## TABLE 5.8 ABOUT HERE

In order to achieve a complete modeling, appropriate marginal distributions must be formally selected once a suitable copula has been chosen. This step can be accomplished using a minimum-distance approach similar to that employed in Section 5.4 for the choice of a parametric elliptical copula. To describe the method briefly, let  $F(y, \beta)$  be a univariate distribution function, where  $\beta \in \mathbb{B}$ , and suppose one wants to test for  $\mathcal{H}_0 : F \in \{F(\cdot, \beta); \beta \in \mathbb{B}\}$  and  $\mathcal{H}_1 : F \notin \{F(\cdot, \beta); \beta \in \mathbb{B}\}$  on the basis of  $Y_1, \dots, Y_n$  i.i.d.  $F$ . First define

$$S_n(\beta) = n \int_{\mathbb{R}} \{F_n(y) - F(y, \beta)\}^2 dF(y, \beta),$$

and consider the goodness-of-fit test statistic and estimator

$$S_n = \inf_{\beta \in \mathbb{B}} S_n(\beta) \quad \text{and} \quad \beta_n = \arg \min_{\beta \in \mathbb{B}} S_n(\beta).$$

The null hypothesis can be rejected whenever  $S_n$  exceeds a critical value estimated from an application of the parametric bootstrap. See Boos (1981) and Parr & Schucany (1982) for more details about this minimum-distance goodness-of-fit method in the univariate case.

Of a particular interest for insurers is the probability that the two claims exceed some large value  $x_0$  simultaneously, *i.e.*  $P(X > x_0, Y > x_0)$ . From

Sklar's Theorem, one can write

$$P(X > x_0, Y > x_0) = 1 - F_X(x_0) - F_Y(x_0) + C\{F_X(x_0), F_Y(x_0)\},$$

where  $F_X, F_Y$  are the marginal distributions. In view of the histograms in the lower panels of Figure 5.2, exponential distributions could possibly be appropriate models for the individual random variables. Considering a bivariate Normal copula  $C_\rho$  for modeling the dependence in the pair, which is radially symmetric, the probability is estimated by

$$P(X > x_0, Y > x_0) = C_{\rho_n}\left(e^{-x_0/\hat{\lambda}_X}, e^{-x_0/\hat{\lambda}_Y}\right),$$

where  $\rho_n = 0.67$ ,  $\hat{\lambda}_X = 2.35$  and  $\hat{\lambda}_Y = 0.86$ . The graphic of this probability as a function of  $x_0$  is presented in Figure 5.3. As one can see, the probability of both claims exceeding five million Euros is very close to 0 under this model. Of course, another choice of a copula and / or margins would have an influence on this probability.

### 5.5.2 The Upper Mississippi river data

The Mississippi river crosses the United States (US) from the North (state of Minnesota) to the South (state of Louisiana), where it discharges water into the Gulf of Mexico. This river is part of the Jefferson–Missouri–Mississippi system that drains about 40% of the US territory. Here, a data set built by Ghizzoni et al. (2012) will be analyzed in the light of our goodness-of-fit methods for the selection of an elliptical copula. The latter is based on a large database ( $n = 22\,600$ ) of simultaneous measures of daily mean discharge

values from a flood event taken at  $d = 18$  stations on the upper part of the Mississippi river from the years 1943 to 2008. These values are actually maximum values observed over certain periods of seven days during the flood event. Afterwards, this database was reduced to only  $n = 89$  independent values. For specific details on the procedure used to select these values see Ghizzoni et al. (2012). The goal of Ghizzoni et al. (2012) was to fit either the Student copula (combined with log-normal and generalized extreme-value marginals) or the skew Student distribution to the  $n = 89$  selected events. However, no formal test was applied in order to validate or discard these models. Using a two-step maximum likelihood approach called *the inference for margins* method (see Joe, 2005), where some chosen parametric marginal distributions are first estimated and then *plugged* into the full likelihood, they obtained  $\hat{\nu} = 9.43$  for the number of degrees of freedom of the Student copula.

The results of the statistical analysis based on the test statistic  $V_n$  are presented in the right part of Table 5.8. Since the estimated correlation matrix  $R_n$  was singular, a modified version was used instead. Indeed, two among the eighteen eigenvalues in the diagonal matrix  $D$ , such that  $R_n = V D V^\top$ , were slightly negative. This problem has been overcome by using a variation of the method proposed by Higham (2002) that consists in replacing  $D$  by  $\tilde{D}$  in which the negative eigenvalues are replaced by some small value, in this case .01. The resulting matrix  $\tilde{R}_n = V \tilde{D} V^\top$  is positive definite and appears to be the closest, with respect to the euclidian norm, to  $R_n$  in the space of positive definite matrices. The results in Table 5.8 are quite congruent with those of Ghizzoni et al. (2012). In fact, all the  $T_\nu$  copulas that were tested

were accepted at the 5% level (except the  $T_{15}$  copula); however, the  $p$ -values are only slightly larger than .05. These findings are also to be shaded by the fact that all the Pearson type II copulas were also accepted. It illustrates the usual difficulty of using goodness-of-fit methods to distinguish between various models in small data sets. Such disappointing results were reported by Genest et al. (2007) for a trivariate data set of size  $n = 47$ . When testing for the hypothesis of a Student copula with an unknown number of degrees of freedom, the minimum-distance statistic yields  $W_n = 7.65$ ,  $p_{n,M}^* = 0.14$  and  $\theta_n = 3.5$ ; for the parametric Pearson type II model,  $W_n = 6.76$ ,  $p_{n,M}^* = 0.06$  and  $\theta_n = 16.00$ . Once again, both models are accepted by a small amount, in conformity with the results for a fixed generator.

### 5.5.3 The Oil currency data

This data set consists of daily log-returns of the oil price, Standard & Poor's 500, and of six currency exchange rates from May 1985 to June 2004. It has been analyzed by Klüppelberg & Kuhn (2009) to illustrate their newly introduced *copula structure analysis*. When considering the last  $n = 904$  observations, these authors concluded to the rejection of the Normal copula and to the acceptance of a Student dependence structure using a goodness-of-fit test introduced by Berg (2009). Based on our results presented on the left part of Table 5.8, the  $T_{15}$  copula is clearly the right choice. Indeed, all other models considered are strongly rejected. When performing the minimum-distance test for a global Student dependence structure, one obtains  $W_n = 0.0264$ ,  $p_{n,M}^* = 1$  and  $\theta_n = 16.00$ . These conclusions agree with that observed

within the graphics of  $\Psi$  versus  $\Psi_{G^2}$ . Those when the null hypothesis is the  $N$ ,  $T_6$ ,  $T_{16}$  and  $Pe_2$  copulas are presented in Figure 5.4.

FIGURE 5.4 ABOUT HERE

#### 5.5.4 The Cook & Johnson data set revisited

As a final illustration, the Uranium exploration data set originally considered by Cook & Johnson (1981, 1986) was analyzed. These  $d = 7$  dimensional data consist of 655 chemical analyses from water samples collected from the Montrose quadrangle of western Colorado (USA). Concentrations were measured for uranium, lithium, cobalt, potassium, caesium, scandium and titanium. A pair-by-pair copula modeling has been considered by Genest et al. (2006) for many Archimedean copula families.

In view of results in the right part of Table 5.8, the Student copulas with  $\nu = 6$  and  $\nu = 9$  degrees of freedom are not rejected at the 5% level, while the other models are clearly rejected. Again, these results are confirmed by the test based on  $W_n$ , where the parametric Student model is accepted ( $W_n = 0.0197$ ,  $p_{n,M}^* = 0.995$ ,  $\theta_n = 7.75$ ) and the Pearson type II copula is rejected ( $W_n = 4.3359$ ,  $p_{n,M}^* < 0.01$ ). A  $T_8$  copula would then be suitable.

#### Acknowledgements

The authors wish to thank Giorgio Roth for providing us with the Upper Mississippi river data set, and Stephan Haug for sending us the Oil currency

data set. Funding in partial support of this work was provided by the Natural Sciences and Engineering Research Council of Canada and by the Fonds Québécois de Recherche Nature et Technologies.



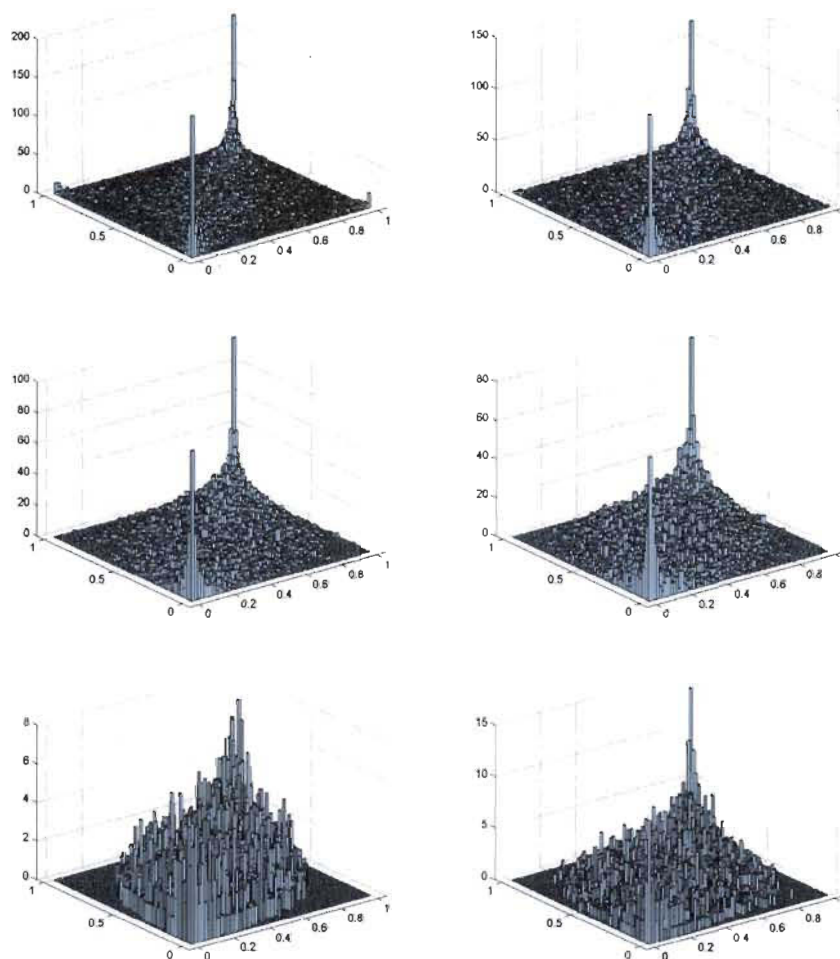


FIGURE 5.1: Three-dimensional histograms of the realizations of 5 000 pairs from the  $T_1$  (upper left panel),  $T_3$  (upper right panel),  $T_9$  (middle left panel), Normal (middle right panel),  $Pe_1$  (lower left panel) and  $Pe_5$  (lower right panel) copulas when  $\tau = .5$ .

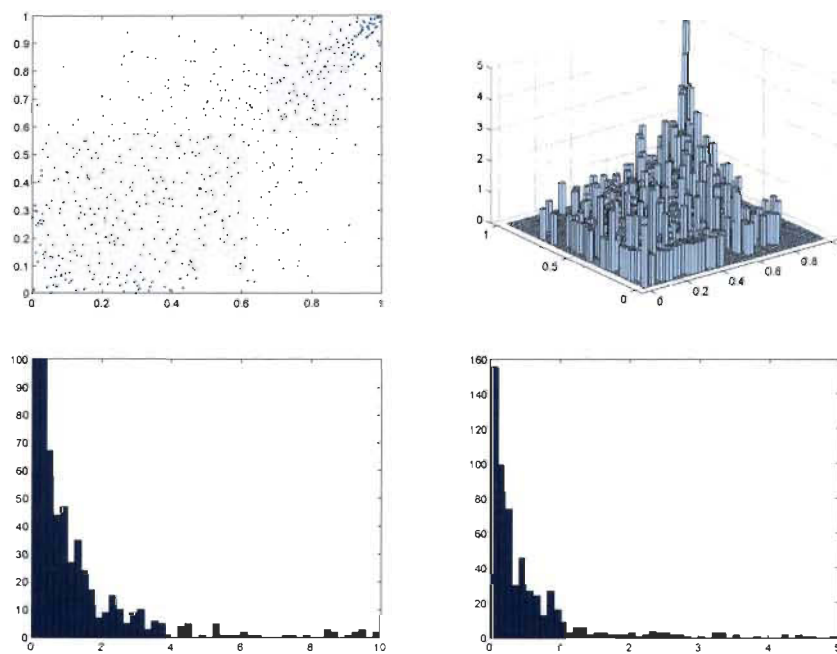


FIGURE 5.2: Scatterplot (upper left panel) and density plot (upper right panel) of the normalized ranks and histograms (lower panels) for the Danish insurance data set.

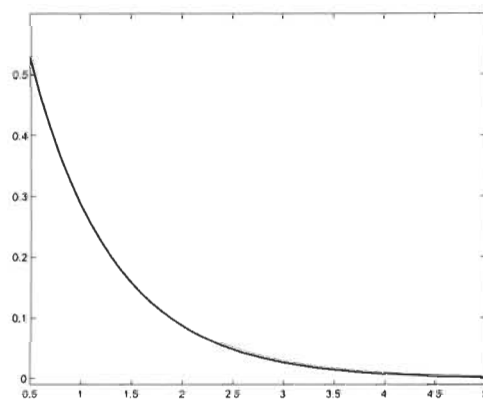


FIGURE 5.3: Probability that both claims in the Danish fire insurance data set exceed  $x_0$  (in millions of Euros).

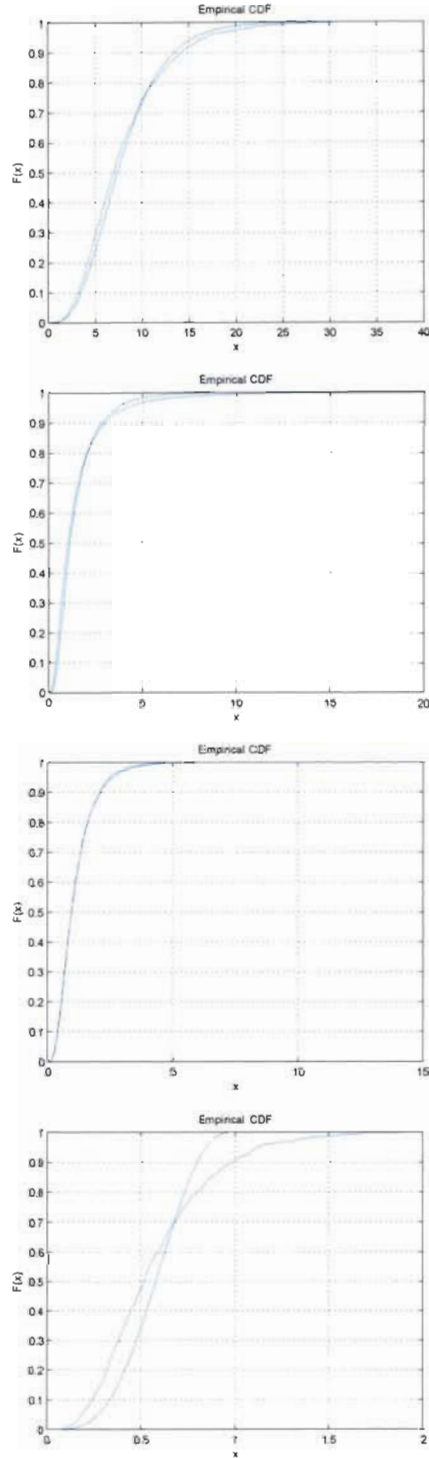


FIGURE 5.4: Plot of  $\Psi_n$  (dashed lines) and  $\Psi_{G^2}$  for the  $N$  (upper left panel),  $T_6$  (upper right panel),  $T_{16}$  (lower left panel) and  $Pe_2$  (lower right panel) elliptical copulas for the Oil currency data set.

TABLE 5.1: Estimation, based on 1 000 replicates, of  $n$  times the mean-squared error  $\widehat{\text{MSE}}_R(R_n)$  of  $R_n$  for estimating the correlation matrix  $R$  in the equicorrelated case

Model	$\rho$	$d = 2$		$d = 3$		$d = 4$		$d = 5$	
		$n = 100$	$n = 250$	$n = 100$	$n = 250$	$n = 100$	$n = 250$	$n = 100$	$n = 250$
$N$	1/4	1.0703	1.0388	0.4800	0.4477	0.3118	0.2739	0.2464	0.2471
	1/2	0.6267	0.6701	0.3920	0.3792	0.2860	0.2897	0.2582	0.2561
	3/4	0.2342	0.2303	0.1517	0.1555	0.1240	0.1362	0.1130	0.1218
$T_1$	1/4	1.7105	1.7624	0.7896	0.7802	0.5111	0.5279	0.3864	0.3918
	1/2	1.2420	1.1652	0.6402	0.6761	0.5114	0.4782	0.4091	0.3926
	3/4	0.5136	0.4710	0.3236	0.3305	0.2633	0.2638	0.2328	0.2321
$T_3$	1/4	1.2011	1.2695	0.5831	0.5946	0.3716	0.3887	0.3179	0.2808
	1/2	0.8690	0.8866	0.4808	0.4804	0.3757	0.3640	0.3184	0.3296
	3/4	0.3356	0.3499	0.2224	0.2143	0.1904	0.1831	0.1544	0.1463
$T_6$	1/4	1.0920	1.1576	0.5013	0.4969	0.3350	0.3560	0.2830	0.2778
	1/2	0.7941	0.7371	0.4605	0.4266	0.3107	0.3270	0.2872	0.3062
	3/4	0.2717	0.2670	0.1742	0.1944	0.1464	0.1568	0.1519	0.1384
$T_9$	1/4	1.0349	1.1108	0.5090	0.5119	0.3330	0.3376	0.2620	0.2718
	1/2	0.7330	0.6892	0.4164	0.4142	0.3161	0.3035	0.2757	0.2602
	3/4	0.2664	0.2666	0.1722	0.1744	0.1468	0.1581	0.1409	0.1372
$\text{Pe}_1$	1/4	0.8081	0.7833	0.3983	0.3792	0.2906	0.2838	0.2189	0.2359
	1/2	0.5037	0.5175	0.3058	0.3027	0.2662	0.2599	0.2224	0.2094
	3/4	0.1864	0.1580	0.1365	0.1260	0.1146	0.1133	0.1051	0.1021
$\text{Pe}_2$	1/4	0.9074	0.8476	0.4072	0.3966	0.2755	0.2744	0.2326	0.2134
	1/2	0.5513	0.5486	0.3335	0.3450	0.2582	0.2664	0.2381	0.2273
	3/4	0.2077	0.1769	0.1343	0.1374	0.1230	0.1059	0.1032	0.1003
$\text{Pe}_5$	1/4	0.9727	0.9236	0.4414	0.4502	0.2666	0.2787	0.2204	0.2458
	1/2	0.6237	0.5920	0.3748	0.3450	0.3120	0.2752	0.2201	0.2439
	3/4	0.2071	0.2044	0.1351	0.1339	0.1293	0.1160	0.1114	0.1067

TABLE 5.2: Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for bivariate meta-elliptical models when  $n = 100$ 

Model under $\mathcal{H}_1$		Model under $\mathcal{H}_0$							
	$\rho$	$T_1$	$T_3$	$T_6$	$T_9$	$N$	$Pe_1$	$Pe_2$	$Pe_5$
$T_1$	1/4	5.9	84.6	98.1	99.0	99.9	100.0	100.0	100.0
	1/2	5.2	80.0	96.2	97.8	99.7	100.0	100.0	99.9
	3/4	5.5	68.8	91.1	92.8	98.2	100.0	99.7	99.2
$T_3$	1/4	75.3	3.0	21.2	31.8	59.7	96.6	90.7	79.7
	1/2	69.3	6.6	21.3	29.7	58.8	95.3	85.3	80.2
	3/4	59.4	5.3	17.4	30.0	54.7	95.1	86.0	75.4
$T_6$	1/4	94.0	8.1	4.7	8.8	23.3	80.1	64.2	42.8
	1/2	92.2	6.7	4.9	6.9	24.9	81.3	61.7	45.0
	3/4	85.5	6.3	5.6	6.4	23.1	79.3	60.8	43.1
$T_9$	1/4	97.2	10.6	4.5	5.1	15.8	72.6	52.6	31.7
	1/2	96.1	10.4	4.4	4.9	14.6	70.4	47.0	27.7
	3/4	93.9	7.0	4.4	5.1	13.0	69.5	47.8	28.3
$N$	1/4	99.5	25.7	8.3	3.8	5.5	44.2	23.7	11.3
	1/2	99.6	25.7	6.0	4.4	4.4	44.5	19.7	10.2
	3/4	98.4	17.0	5.7	5.1	4.9	43.7	18.5	10.8
$Pe_1$	1/4	100.0	81.1	45.3	34.3	14.8	4.8	5.8	7.6
	1/2	100.0	79.0	40.8	29.6	14.3	5.8	6.3	6.6
	3/4	100.0	80.1	37.0	29.5	14.4	4.4	6.4	7.5
$Pe_2$	1/4	100.0	62.7	25.4	15.9	7.2	10.7	6.0	3.1
	1/2	100.0	64.5	24.3	17.5	6.8	10.7	4.2	6.4
	3/4	100.0	59.0	21.4	15.6	6.9	10.2	4.9	3.4
$Pe_5$	1/4	99.9	41.6	13.9	7.7	4.4	22.7	11.8	5.2
	1/2	99.8	42.6	12.8	7.1	4.4	22.3	9.4	7.1
	3/4	100.0	39.5	11.5	7.3	4.4	24.5	8.4	5.1
CL	1/4	99.5	19.7	5.2	4.3	6.4	50.3	31.8	13.7
	1/2	98.9	18.0	3.1	5.1	8.3	57.8	36.2	20.1
	3/4	97.9	13.2	4.4	4.2	10.4	63.4	37.7	22.9
GH	1/4	98.3	11.5	3.5	4.1	12.8	68.4	43.9	26.9
	1/2	91.5	6.2	4.2	6.1	17.4	76.2	53.3	35.6
	3/4	57.3	6.2	4.1	7.8	12.8	76.1	53.5	39.7

TABLE 5.3: Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for bivariate meta-elliptical models when  $n = 250$ 

Model under $\mathcal{H}_1$	$\rho$	Model under $\mathcal{H}_0$							
		$T_1$	$T_3$	$T_6$	$T_9$	$N$	$Pe_1$	$Pe_2$	$Pe_5$
$T_1$	1/4	5.8	99.8	100.0	100.0	100.0	100.0	100.0	100.0
	1/2	4.4	99.2	100.0	100.0	100.0	100.0	100.0	100.0
	3/4	3.6	96.5	99.7	100.0	100.0	100.0	100.0	100.0
$T_3$	1/4	99.7	4.4	39.6	56.7	92.7	100.0	99.9	97.6
	1/2	99.3	4.7	40.0	59.7	91.3	100.0	99.7	97.5
	3/4	97.1	4.2	35.4	53.8	86.4	100.0	99.8	96.8
$T_6$	1/4	100.0	18.0	4.6	9.2	39.8	98.8	93.3	68.5
	1/2	100.0	17.0	5.2	10.9	41.0	99.6	92.2	69.4
	3/4	100.0	16.5	4.8	11.1	37.9	99.1	91.2	67.1
$T_9$	1/4	100.0	36.0	3.7	5.0	22.5	95.7	80.5	48.2
	1/2	100.0	32.1	5.6	3.7	27.3	96.5	77.5	46.6
	3/4	100.0	31.4	6.0	5.1	21.5	97.6	81.0	46.6
$N$	1/4	100.0	78.9	17.2	9.6	3.4	72.6	35.4	10.4
	1/2	100.0	76.7	21.2	9.0	5.5	76.0	33.1	11.8
	3/4	100.0	73.9	19.4	11.0	4.4	75.8	35.2	10.4
$Pe_1$	1/4	100.0	100.0	97.1	90.6	51.0	4.0	5.4	19.3
	1/2	100.0	100.0	98.5	92.4	56.8	2.8	4.5	23.0
	3/4	100.0	100.0	98.4	91.2	57.4	2.5	5.9	21.7
$Pe_2$	1/4	100.0	99.5	80.6	62.5	17.8	9.6	2.4	5.8
	1/2	100.0	99.6	83.0	63.8	20.7	9.3	2.2	5.5
	3/4	100.0	99.8	85.8	65.3	22.1	9.9	2.8	5.2
$Pe_5$	1/4	100.0	95.3	50.0	31.8	6.3	39.5	8.9	3.5
	1/2	100.0	95.8	49.7	25.7	6.6	36.3	9.4	3.4
	3/4	100.0	96.4	48.9	27.3	6.6	33.7	7.8	3.3
CL	1/4	100.0	64.1	13.4	5.2	5.2	85.0	49.5	16.5
	1/2	100.0	53.9	10.2	5.4	8.9	90.3	61.3	25.0
	3/4	100.0	47.0	8.8	4.4	12.9	93.5	66.3	33.1
GH	1/4	100.0	39.9	5.2	3.5	19.0	94.9	76.5	39.8
	1/2	100.0	20.1	3.8	8.4	28.1	99.2	86.4	57.3
	3/4	99.9	10.7	4.1	8.1	33.3	98.6	91.7	70.2

TABLE 5.4: Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for meta-elliptical models in the equicorrelated case for  $d \in \{3, 4\}$  and  $n = 100$

Model under $\mathcal{H}_1$	$\rho$	$d = 3$								$d = 4$							
		Model under $\mathcal{H}_0$								Model under $\mathcal{H}_0$							
		$T_1$	$T_3$	$T_6$	$T_9$	$N$	$Pc_1$	$Pc_2$	$Pe_5$	$T_1$	$T_3$	$T_6$	$T_9$	$N$	$Pc_1$	$Pc_2$	$Pe_5$
$T_1$	1/4	3.8	97.0	100.0	100.0	100.0	100.0	100.0	100.0	5.4	99.3	100.0	100.0	100.0	100.0	100.0	100.0
	1/2	4.5	95.2	100.0	100.0	100.0	100.0	100.0	100.0	4.0	98.7	100.0	100.0	100.0	100.0	100.0	100.0
	3/4	4.2	88.3	98.6	99.6	100.0	100.0	100.0	100.0	3.2	94.5	99.7	100.0	100.0	100.0	100.0	100.0
$T_3$	1/4	98.3	5.1	37.8	58.2	91.0	100.0	99.9	98.1	99.8	4.9	49.8	72.8	98.3	100.0	100.0	99.9
	1/2	96.7	3.7	38.7	55.4	90.4	99.9	99.7	98.0	99.8	3.3	46.0	70.8	98.0	100.0	100.0	99.9
	3/4	93.3	4.1	26.2	46.8	84.6	99.9	99.8	97.4	99.2	4.3	37.0	61.9	95.7	100.0	100.0	99.6
$T_6$	1/4	100.0	16.6	4.9	12.1	48.0	98.5	95.5	82.3	100.0	29.8	5.2	13.5	69.0	99.7	98.7	93.3
	1/2	99.9	17.0	5.3	11.0	46.6	96.7	94.3	81.7	100.0	30.8	6.0	13.0	66.8	99.5	99.3	94.2
	3/4	99.7	14.8	4.8	10.7	44.8	97.2	91.9	77.0	100.0	30.7	4.8	10.5	58.7	99.8	98.6	89.4
$T_9$	1/4	100.0	30.8	5.6	6.0	27.3	94.4	86.3	60.9	100.0	61.2	6.5	4.4	12.3	98.1	94.4	78.2
	1/2	100.0	32.7	4.1	5.1	26.0	93.2	87.3	62.6	100.0	57.0	5.2	4.4	41.2	98.3	95.7	76.2
	3/4	100.0	27.8	5.0	6.1	23.9	93.8	82.2	57.2	100.0	55.8	8.2	4.7	33.0	98.0	95.2	72.5
$N$	1/4	100.0	75.1	23.6	10.2	5.0	63.6	44.0	17.5	100.0	98.2	47.5	16.7	5.4	72.3	50.7	19.8
	1/2	100.0	76.9	19.9	7.8	5.4	65.9	44.3	19.5	100.0	97.0	44.9	18.8	6.4	74.5	57.9	18.2
	3/4	100.0	68.9	19.0	10.4	3.7	65.7	38.8	16.4	100.0	96.6	43.5	17.9	5.2	74.6	54.0	20.7
$Pc_1$	1/4	100.0	100.0	97.5	89.2	51.0	5.8	9.9	22.4	100.0	100.0	99.9	99.5	77.7	4.2	14.1	43.1
	1/2	100.0	99.8	96.6	85.7	46.8	5.6	8.7	20.4	100.0	100.0	100.0	99.7	82.6	3.7	13.9	37.0
	3/4	100.0	100.0	95.5	84.3	47.1	4.8	9.4	21.6	100.0	100.0	99.9	99.5	77.4	5.9	10.4	33.3
$Pc_2$	1/4	100.0	99.6	82.7	65.2	21.6	8.5	5.6	6.5	100.0	100.0	98.8	93.2	41.1	4.8	4.1	14.2
	1/2	100.0	99.8	80.8	62.5	23.2	7.8	5.4	8.9	100.0	100.0	98.9	92.5	44.1	5.6	4.8	12.6
	3/4	100.0	99.6	82.3	59.9	21.6	6.2	5.5	7.5	100.0	100.0	98.5	92.8	43.8	5.5	5.3	12.4
$Pe_5$	1/4	100.0	95.5	53.3	29.2	5.8	28.2	14.1	6.0	100.0	100.0	88.1	65.0	13.2	24.7	11.8	4.5
	1/2	100.0	95.6	52.1	30.1	6.9	28.7	12.8	6.2	100.0	100.0	88.7	60.8	12.2	26.0	14.1	4.5
	3/4	100.0	95.3	50.4	28.0	6.3	26.4	11.9	5.3	100.0	99.9	88.0	61.5	13.3	26.0	11.2	5.4



TABLE 5.5: Percentages of rejection, as estimated from 1 000 replicates, of the goodness-of-fit tests for the parametric meta-elliptical models  $T$  and  $Pe$  in the equicorrelated case with  $\rho = 1/2$  for  $d \in \{2, 3\}$  when  $n \in \{50, 100\}$

Model under $\mathcal{H}_1$	$n = 50$				$n = 100$			
	$d = 2$		$d = 3$		$d = 2$		$d = 3$	
	Model under $\mathcal{H}_0$		Model under $\mathcal{H}_0$		Model under $\mathcal{H}_0$		Model under $\mathcal{H}_0$	
	$T$	$Pe$	$T$	$Pe$	$T$	$Pe$	$T$	$Pe$
$T_1$	6.0	95.8	6.0	98.7	4.7	99.5	6.1	100.0
$T_3$	4.2	58.1	6.6	66.2	1.7	71.1	5.0	96.7
$T_6$	4.7	35.4	8.8	34.2	3.3	41.4	4.6	70.6
$T_9$	5.7	25.3	7.7	23.3	3.8	25.8	3.5	54.6
$N$	6.2	13.1	11.2	7.4	4.9	9.9	11.2	11.8
$Pe_1$	15.3	8.0	54.1	16.8	33.4	6.4	88.9	9.1
$Pe_2$	11.4	7.7	32.4	8.7	19.5	4.1	62.5	3.9
$Pe_5$	7.9	8.0	19.2	4.8	9.8	4.4	29.8	6.0

TABLE 5.6: Estimation, based on 1 000 replicates, of  $n$  times the mean-squared errors of  $\hat{\theta}_n$  for estimating the parameter of the elliptical generator in the Student and Pearson type II copulas when  $n = 50$

Model under $\mathcal{H}_1$	Dimension $d$		
	2	3	4
$T_1$	14.64	9.61	11.68
$T_3$	117.23	81.26	72.01
$T_6$	343.10	289.85	246.93
$T_9$	414.07	375.83	364.22
$\text{Pe}_1$	229.42	262.74	212.15
$\text{Pe}_2$	201.63	161.40	88.21
$\text{Pe}_5$	160.22	83.39	387.26

TABLE 5.7: Results of the goodness-of-fit tests on the Danish fire insurance and Upper Mississippi river data sets for the null hypothesis of a meta-elliptical distribution with a fixed generator ( $p$ -values and critical values estimated from  $M = 1\,000$  parametric bootstrap samples)

Model under $\mathcal{H}_0$	Danish fire insurance data			Upper Mississippi river data		
	$V_n$	$p$ -value	Critical value	$V_n$	$p$ -value	Critical value
$T_1$	1.7674	0.01	0.1228	13.5651	0.0520	14.1875
$T_3$	0.2074	0.0040	0.1252	9.2496	0.0800	11.8633
$T_5$	0.0400	0.7300	0.1178	7.6496	0.1040	13.7071
$T_9$	0.0220	0.9760	0.1224	7.1185	0.1040	10.5601
$T_{15}$	0.0233	0.9760	0.1149	6.7834	0.0400	6.2213
$N$	0.0516	0.5810	0.1176	6.7896	0.0760	8.3766
$Pe_1$	0.4882	0.01	0.1223	9.9542	0.0680	11.8050
$Pe_2$	0.2853	0.01	0.1235	8.4941	0.1080	11.8820
$Pe_5$	0.1370	0.0270	0.1176	7.7326	0.0560	9.6842
$Pe_{10}$	0.0866	0.1960	0.1240	7.9703	0.0800	9.4569

TABLE 5.8: Results of the goodness-of-fit tests on the Oil currency and Uranium exploration data sets for the null hypothesis of a meta-elliptical distribution with a fixed generator ( $p$ -values and critical values estimated from  $M = 250$  parametric bootstrap samples)

Model under $\mathcal{H}_0$	Oil currency data			Uranium exploration data		
	$V_n$	$p$ -value	Critical value	$V_n$	$p$ -value	Critical value
$T_1$	45.1691	< 0.01	0.2922	22.1950	< 0.01	0.3361
$T_3$	8.6763	< 0.01	0.2480	2.7615	< 0.01	0.2225
$T_6$	1.7985	< 0.01	0.2127	0.1422	0.1810	0.2052
$T_9$	0.4727	< 0.01	0.2179	0.0526	0.8670	0.2029
$T_{15}$	0.0308	1.0000	0.2368	0.4561	< 0.01	0.2256
$N$	1.0803	< 0.01	0.3109	2.2810	< 0.01	0.2617
$Pe_1$	14.0350	< 0.01	0.9329	13.3178	< 0.01	0.7454
$Pe_2$	9.7005	< 0.01	0.7094	9.8694	< 0.01	0.5670
$Pe_5$	5.1667	< 0.01	0.5022	6.0796	< 0.01	0.4297
$Pe_{10}$	3.1627	< 0.01	0.3912	4.3359	< 0.01	0.3485

## CHAPITRE 6

## CONCLUSION

L'objectif principal de ce mémoire était de développer et d'étudier de nouvelles procédures d'adéquation pour les lois méta-elliptiques. Dans un premier temps, on a considéré la représentation d'un vecteur  $\mathbf{X}$  de loi elliptique sous la forme  $\mathbf{X} \stackrel{d}{=} \mu + \mathcal{G} A \mathcal{U} (*)$ , où  $\mu$  est un vecteur moyenne,  $\mathcal{G}$  est une variable aléatoire positive,  $A$  est une matrice telle que  $A^\top A = \Sigma$  est définie positive et  $\mathcal{U}$  est un vecteur uniformément distribué sur la sphère unité dans  $\mathbb{R}^d$ . On note alors  $\mathbf{X} \sim \mathcal{E}(\mu, \Sigma, g)$ , où  $g : [0, \infty] \rightarrow [0, \infty)$  caractérise la loi de  $\mathcal{G}$ .

Par la suite, on a considéré plus généralement les vecteurs aléatoires dont la *structure de dépendance*, c'est-à-dire la copule, est extraite d'une loi elliptique. On dit d'un tel vecteur aléatoire  $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$  qu'il suit une loi méta-elliptique de marges  $\mathbf{F} = (F_1, \dots, F_d)^\top$  si

$$\mathbf{X} = (Q_g^{-1} \circ F_1(Y_1), \dots, Q_g^{-1} \circ F_d(Y_d))^\top \sim \mathcal{E}(\mathbf{0}, \Sigma, g),$$

où  $Q_g$  est la fonction de répartition univariée. On note  $\mathbf{Y} \sim \mathcal{ME}(\Sigma, g, \mathbf{F})$ .

Dans la représentation stochastique d'un vecteur de loi elliptique, c'est la variable aléatoire  $\mathcal{G}$  qui caractérise entièrement la distribution. En effet,

les autres composantes sont un paramètre de lieu, soit  $\mu$ , et un paramètre d'échelle, c'est-à-dire  $A$ . Ainsi, si on veut vérifier l'adéquation d'un jeu de données à une loi elliptique (ou éventuellement *méta-elliptique*), on peut se concentrer uniquement sur  $\mathcal{G}$ . En considérant de nouveau la représentation (\*), on déduit aisément que  $\mathcal{G}^2 = (\mathbf{X} - \mu)^\top \Sigma^{-1} (\mathbf{X} - \mu)$ . Cette constatation est à la base des méthodes statistiques développées dans ce mémoire.

Supposons que l'on observe un échantillon  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  dans  $\mathbb{R}^d$ , où chaque composante est une *copie* d'un vecteur aléatoire  $\mathbf{Y}$ . Afin de tester l'hypothèse que  $\mathbf{Y} \sim \mathcal{ME}(\Sigma, g, \mathbf{F})$ , où  $\Sigma$  et  $\mathbf{F}$  sont inconnus, on a d'abord construit le *pseudo-échantillon*  $\mathbf{X}_{1,n}, \dots, \mathbf{X}_{n,n}$  tel que pour tout  $i \in \{1, \dots, n\}$ ,

$$\mathbf{X}_{i,n} = (Q_g^{-1} \circ F_{n1}(Y_1), \dots, Q_g^{-1} \circ F_{nd}(Y_d))^\top,$$

où  $\mathbf{F}_n = (F_{n1}, \dots, F_{nd})^\top$  est le vecteur des fonctions de répartition marginales empiriques. Ensuite, des *copies*  $\mathcal{G}_{1,n}^2, \dots, \mathcal{G}_{n,n}^2$  de la variable aléatoire  $\mathcal{G}^2$  sont obtenues via  $\mathcal{G}_{i,n}^2 = \mathbf{X}_{i,n}^\top \Sigma_n^{-1} \mathbf{X}_{i,n}$ , où  $\Sigma_n$  est une estimation de la matrice  $\Sigma$ . Ensuite, on a comparé la fonction de répartition empirique  $\Psi_n$  de  $\mathcal{G}_{1,n}^2, \dots, \mathcal{G}_{n,n}^2$  avec celle de  $\mathcal{G}^2$  attendue sous l'hypothèse nulle, notée  $\Psi_{\mathcal{G}^2}$ . Ceci a amené à définir une statistique de test basée sur la distance de Cramér-von Mises entre ces deux fonctions de répartition. Le calcul des  $p$ -valeurs a été effectué à l'aide d'une adaptation de la méthode du bootstrap paramétrique.

Les outils décrits précédemment ont été généralisés aux cas où la variable aléatoire  $\mathcal{G}^2$  suit une loi avec un paramètre inconnu. Afin d'en tenir compte, une approche dite à *distance minimale* a été développée. Celle-ci permet à la fois d'estimer le paramètre et de tester l'adéquation à un certain modèle paramétrique méta-elliptique. On peut donc, avec cette méthode, vérifier

par exemple que des données proviennent d'une loi dont la dépendance est de type Student avec un degré de liberté quelconque.

Une étude de simulation Monte–Carlo à grande échelle a montré que les nouvelles méthodologies proposées sont très puissantes. En effet, les tests statistiques permettent de sélectionner un modèle méta-elliptique adéquat avec une très grande probabilité de succès. On a aussi illustré l'utilité de nos méthodes sur de vraies jeux de données. Ainsi, on a modélisé la dépendance, via une loi méta-elliptique, de réclamations d'assurance suite à un incendie, des débits de la rivière Mississippi, mesurées à partir de 18 stations différentes, du cours du pétrole et de devises de quelques pays, ainsi que du célèbre jeu de données de Cook & Johnson (1981) concernant la concentration d'éléments chimiques dans des échantillons d'eau.

Plusieurs avenues de recherche intéressantes pourraient découler de ce travail de maîtrise. D'abord, bien que nos nouvelles méthodologies se sont avérées très efficaces, la convergence de la fonction empirique  $\Psi_n$  n'a pas été démontrée formellement. Ainsi, pour établir rigoureusement que la méthode est valide, il faudrait fournir la preuve que le *processus empirique*  $\sqrt{n}\{\Psi_n(\gamma) - \Psi_{G^2}(\gamma)\}$  converge en loi vers une certaine limite gaussienne. Ce problème est extrêmement compliqué et fait appel à la théorie des pseudo-observations développée dans les travaux de Ghoudi & Rémillard (1998) et Ghoudi & Rémillard (2004). Une autre voie prometteuse concerne l'adaptation de nos nouveaux outils d'inférence dans des plans d'expérience non standards. On peut penser aux modèles flexibles issus de la *pair-copula construction* introduite par Aas et al. (2009), ou encore aux distributions hiérarchiques.

# Bibliographie

- AAS, K., CZADO, C., FRIGESSI, A. & BAKKEN, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44**, 182–198.
- ABDOUS, B., GENEST, C. & RÉMILLARD, B. (2005). Dependence properties of meta-elliptical distributions. In *Statistical modeling and analysis for complex data problems*, vol. 1 of *GERAD 25th Anniv. Ser.* New York : Springer, pp. 1–15.
- BARBE, P., GENEST, C., GHOUDI, K. & RÉMILLARD, B. (1996). On Kendall's process. *J. Multivariate Anal.* **58**, 197–229.
- BERG, D. (2009). Copula goodness-of-fit testing : an overview and power comparison. *Eur. J. Financ.* **15**, 675–701.
- BERG, D. & QUESSY, J.-F. (2009). Local power analyses of goodness-of-fit tests for copulas. *Scand. J. Stat.* **36**, 389–412.
- BILODEAU, M. & BRENNER, D. (1999). *Theory of multivariate statistics*. Springer Texts in Statistics. New York : Springer-Verlag.



- BOOS, D. D. (1981). Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc.* **76**, 663–670.
- CAMBANIS, S., HUANG, S. & SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11**, 368–385.
- COOK, R. D. & JOHNSON, M. E. (1981). A family of distributions for modelling nonelliptically symmetric multivariate data. *J. Roy. Statist. Soc. Ser. B* **43**, 210–218.
- COOK, R. D. & JOHNSON, M. E. (1986). Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics* **28**, 123–131.
- DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.
- ELTOFT, T., KIM, T. & LEE, T. (2006). On the multivariate Laplace distribution. *Signal Processing Letters, IEEE* **13**, 300–303.
- EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T. (1997). *Modelling extremal events*, vol. 33 of *Applications of Mathematics (New York)*. Berlin : Springer-Verlag. For insurance and finance.
- FANG, H.-B., FANG, K.-T. & KOTZ, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* **82**, 1–16.
- FANG, K. T., KOTZ, S. & NG, K. W. (1990). *Symmetric multivariate and related distributions*, vol. 36 of *Monographs on Statistics and Applied Probability*. London : Chapman and Hall Ltd.

- FANG, K. T. & ZHANG, Y. T. (1990). *Generalized multivariate analysis*. Berlin : Springer-Verlag.
- GENEST, C., FAVRE, A., BÉLIVEAU, J. & JACQUES, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research* **43**, W09401.
- GENEST, C., GHOUDI, K. & RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.
- GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.* **33**, 337–366.
- GENEST, C. & RÉMILLARD, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. Henri Poincaré Probab. Stat.* **44**, 1096–1127.
- GENEST, C., RÉMILLARD, B. & BEAUDOIN, D. (2009). Goodness-of-fit tests for copulas : a review and a power study. *Insurance Math. Econom.* **44**, 199–213.
- GENEST, C. & RIVEST, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.* **88**, 1034–1043.
- GENEST, C. & RIVEST, L.-P. (2001). On the multivariate probability integral transformation. *Statist. Probab. Lett.* **53**, 391–399.

- GHIZZONI, T., ROTH, G. & RUDARI, R. (2012). Multisite flooding hazard assessment in the upper mississippi river. *J. Hydrol.* **412–413**, 101–113.
- GHOUDI, K. & RÉMILLARD, B. (1998). Empirical processes based on pseudo-observations. In *Asymptotic methods in probability and statistics (Ottawa, ON, 1997)*. Amsterdam : North-Holland, pp. 171–197.
- GHOUDI, K. & RÉMILLARD, B. (2004). Empirical processes based on pseudo-observations. II. The multivariate case. In *Asymptotic methods in stochastics*, vol. 44 of *Fields Inst. Commun.* Providence, RI : Amer. Math. Soc., pp. 381–406.
- HIGHAM, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.* **22**, 329–343.
- JOE, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.* **94**, 401–419.
- KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32**, 419–438.
- KELLY, K. & KRZYSZTOFOWICZ, R. (1997). A bivariate meta-gaussian density for use in hydrology. *Stoch. Hydrol. Hydraul.* **11**, 17–31.
- KLÜPPELBERG, C. & KUHN, G. (2009). Copula structure analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71**, 737–753.
- KOTZ, S. (1975). Multivariate distributions at a cross-road, gp patil, s. kotz, jk ord, editors statistical distributions in scientific work, vol. 1 d.

- LANDSMAN, Z. (2009). Elliptical families and copulas : tilting and premium ; capital allocation. *Scand. Actuar. J.* , 85–103.
- LANDSMAN, Z. M. & VALDEZ, E. A. (2003). Tail conditional expectations for elliptical distributions. *N. Am. Actuar. J.* **7**, 55–71.
- LEE, A. J. (1990). *U-statistics*, vol. 110 of *Statistics : Textbooks and Monographs*. New York : Marcel Dekker Inc. Theory and practice.
- LINDSKOG, F., MCNEIL, A. & SCHMOCK, U. (2003). Kendall's tau for elliptical distributions. *Credit risk : Measurement, evaluation and management* , 149–156.
- MALEVERGNE, Y. & SORNETTE, D. (2003). Testing the Gaussian copula hypothesis for financial assets dependences. *Quant. Finance* **3**, 231–250.
- MESFIOUI, M., QUESSY, J.-F. & TOUPIN, M.-H. (2009). On a new goodness-of-fit process for families of copulas. *Canad. J. Statist.* **37**, 80–101.
- NELSEN, R. B. (2006). *An introduction to copulas*, 2nd edition. New York : Springer.
- OWEN, J. & RABINOVITCH, R. (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *J. Financ.* , 745–752.
- PARR, W. C. & SCHUCANY, W. R. (1982). Minimum distance estimation and components of goodness-of-fit statistics. *J. Roy. Statist. Soc. Ser. B* **44**, 178–189.

- POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27**, 43–70.
- QUESSY, J.-F. (2005). *Théorie et application des copules : tests d'adéquation, tests d'indépendance, et bornes pour la valeur-à-risque*. Ph.D. thesis, Université Laval, Québec, Canada.
- RYTGAARD, M. (1996). Simulation experiments on the mean residual life function  $m(x)$ . In *Proceedings of the XXVII ASTIN Colloquium, Copenhagen, Denmark*, vol. 1.
- SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *Ann. of Math. (2)* **39**, 811–841.
- SKLAR, M. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.
- TSUKAHARA, H. (2005). Semiparametric estimation in copula models. *Canadian J. Statist.* **33**, 357–375.
- WANG, X., GEBREMICHAEL, M. & YAN, J. (2010). Weighted likelihood copula modeling of extreme rainfall events in connecticut. *J. Hydrol.* **390**, 108–115.

# ANNEXE A

## PROGRAMMES EN MATLAB

### A.1 Simulation de variables aléatoires

#### A.1.1 Quelques lois univariées

```
1 function X = sim_univ_elliptical(law,beta,n)
2
3 % Gamma : law = 1 ;
4 % Fisher : law = 2 ;
5 % Weibull : law = 3 ;
6 % Rayleigh : law = 4 ;
7 % Log-Normal : law = 5
8
9 if (law==1)
10     X = gamrnd(beta(1),beta(2),n,1);
11 elseif (law==2)
12     X = frnd(beta(1),beta(2),n,1);
13 elseif (law==3)
14     X = wblrnd(beta(1),beta(2),n,1);
15 elseif (law==4)
16     X = raylrnd(beta(1),n,1);
17 elseif (law==5)
18     X = lognrnd(beta(1),beta(2),n,1);
19 end
```

### A.1.2 Quelques lois multivariées

```

1 function X = sim_elliptical(R,model,n)
2
3 %Generates multivariate random variables
4 s = size(R);
5 d = s(1);
6 X = zeros(n,d);
7
8 law = model(1);
9
10 if (law==1) %Normal copula
11     mu = zeros(1,d);
12     X = mvnrnd(mu,R,n);
13 elseif (law==2) %Student copula
14     theta = model(2);
15     X = mvtrnd(R,theta,n);
16 elseif (law==3) %Pearson type II copula
17     A = chol(R).';
18     theta = model(2);
19     for i=1:n
20         U = normrnd(0,1,d,1);
21         U = U/norm(U);
22         G = sqrt(betarnd(d/2,theta+1));
23         X(i,:) = (G*A*U)';
24     end
25 elseif (law==4) %Exponential-power copula
26     A = chol(R).';
27     theta = model(2);
28     for i=1:n
29         U = normrnd(0,1,d,1);
30         U = U/norm(U);
31         G = sqrt(gamrnd(d/2,2/theta));
32         X(i,:) = (G*A*U)';
33     end
34 elseif (law==5) %Clayton copula
35     tau = model(2);
36     X = sim_copula(n,1,tau);
37 elseif (law==6) %Gumbel-Hougaard copula
38     tau = model(2);
39     X = sim_copula(n,4,tau);
40 end

```

## A.2 Procédures utilitaires

### A.2.1 Estimation de la matrice de corrélation

```

1 function R = EstimationR(X)
2
3 %Estimation of the correlation matrix R in meta-elliptical
4 %models based on the inversion of Kendall's tau
5
6 s = size(X); d = s(2);
7 R = ones(d,d);
8
9 for i=1:d-1
10     for j=i+1:d
11         R(i,j) = sin((pi/2)*tau([X(:,i),X(:,j)])));
12         R(j,i) = R(i,j);
13     end
14 end

```

### A.2.2 Densité de la loi Pearson type II

```

1 function q = q_PearsonII(x,theta,d)
2
3 %density function of the Pearson type II distribution
4 q = gamma(theta + d/2 + 1)/(gamma(theta + (d+1)/2)*gamma(1/2))
5     *(1 - x^2)^(theta + (d-1)/2);

```

### A.2.3 Fonction de répartition Pearson II inverse

```

1 function x = Q_PearsonII_Inv(u,theta,d,N)
2
3 %Numerical computation of the inverse marginal
4 %of the elliptical Pearson type II distribution
5 %with parameter theta and in dimension d
6
7 Q = 0;
8 x = -1;
9
10 while (Q<=u)
11     Q = Q + q_PearsonII(x + (1/(2*N)),theta,d)*(1/N);
12     x = x + 1/N;
13 end
14 x = x - 1/(2*N);

```



## A.3 Adéquation à une loi méta-elliptique de générateur fixe

### A.3.1 Fonction de répartition de la variable aléatoire $\mathcal{G}^2$

```
1 function g = cdf_G2(r,model)
2
3 % Loi normale : model=(1,d)
4 % Loi Student : model=(2,ddl,d)
5 % Loi PearsonII : model = (3,theta,d)
6 % Loi Exp Power : model = (4,theta,d)
7
8 if (model(1) == 1)
9     d = model(3);
10    g = chi2cdf(r,d);
11 elseif ( model(1) == 2 )
12    ddl = model(2);
13    d = model(3);
14    g = fcdf(r/d,d,ddl);
15 elseif ( model(1) == 3)
16    theta = model(2);
17    d = model(3);
18    g = betacdf(r,d/2,theta+1);
19 elseif ( model(1) == 4)
20    theta = model(2);
21    d = model(3);
22    g = gamcdf(r,d/2,2/theta);
23 end
```

### A.3.2 Calcul de la statistique de test $V_n$

```

1 function [Vn,Rn] = Vn_elliptical(Y,model)
2
3 % Returns the test statistic Vn under a given model
4 % and the estimated correlation matrix
5 %
6 % Normal : model=1
7 % Student : model=(2,theta)
8 % PearsonII : model=(3,theta)
9
10 s = size(Y); n = s(1); d = s(2);
11
12 law = model(1);
13
14 % Estimation of the correlation matrix
15 Rn = EstimationR(Y);
16
17 % Transformation into normalized ranks
18 RANKS = zeros(n,d);
19 for i=1:d
20     RANKS(:,i) = rankit(Y(:,i)) / (n+1);
21 end
22
23 % Construction of the vectors X1, ..., Xn
24 if (law==1)
25     X = norminv(RANKS,0,1);
26 elseif (law==2)
27     theta = model(2);
28     X = tinv(RANKS,theta);
29 elseif (law==3);
30     X = zeros(n,d);
31     theta = model(2);
32     for i=1:n
33         for j=1:d
34             X(i,j) = Q_PearsonII_Inv(RANKS(i,j),theta,d,n+1);
35         end
36     end
37 end
38
39 % Construction of the pseudo-observations  $G_1^2, \dots, G_n^2$ 
40 G = zeros(1,n);
41 for i=1:n
42     G(i) = ( X(i,:) / Rn ) * X(i,:)' ;
43 end
44
45 % Computation of the test statistic
46 A1 = 0; A2 = 0;
47 for i=1:n
48     A2 = A2 + ( 1 - (cdf_G2(G(i),[model,d]))^2 );
49     for j=1:n
50         A1 = A1 + ( 1 - cdf_G2(max(G(i),G(j)),[model,d]));

```

```

51     end
52 end
53 Vn = (A1/n) - A2 + (n/3);

```

### A.3.3 Exécution du test basé sur $V_n$

```

1  function [Vn,PV,CV, test] = GoF_Vn_elliptical(Y,models,M)
2
3  % 'models' is a K x 2 matrix where each line
4  % is the elliptical copula to be tested
5
6  temp = size(models);
7  J = temp(1);
8
9  Vn = zeros(J,1);
10 PV = zeros(J,1);
11 CV = zeros(J,1);
12 test = zeros(J,1);
13
14 for j=1:J
15     model = models(j,:);
16     % Value of the test statistic Vn
17     % and estimated correlation matrix
18     [Vn(j),Rn] = Vn_elliptical(Y,model);
19
20     % Parametric bootstrap
21     Vn_boot = zeros(M,1);
22     n = 100;
23     parfor h=1:M
24         Y_boot = sim_elliptical(Rn,model,n);
25         [Vn_boot(h),~] = Vn_elliptical(Y_boot,model);
26     end
27
28     % Estimated critical value
29     CV(j) = prctile(Vn_boot,95);
30
31     % Estimated p-value
32     PV(j) = sum(Vn_boot>Vn(j)) / M;
33
34     if (PV(j) < 0.05)
35         test(j) = 1;
36     end
37 end

```

## A.4 Adéquation à une loi méta-elliptique de générateur paramétrique

### A.4.1 Calcul de la statistique de test $W_n$

```

1  function [Wn,Rn,theta_n] = Wn_elliptical(Y,law,a,b,PAS)
2
3  %Student : law = 2 ;
4  %PearsonIII : law = 3 ;
5
6  if (law==2)
7      a = max(1,a);
8  elseif (law==3)
9      a = max(-.5,a);
10 end
11
12 %Calcul de la statistique de test
13 T = floor((b-a)/PAS) + 1;
14 Wn_theta = zeros(T,1);
15
16 theta = a;
17 for i=1:T
18     [Wn_theta(i),Rn] = Vn_elliptical(Y,[law,theta]);
19     theta = theta + PAS;
20 end
21
22 % Value of the minimum-distance statistic and the estimator
23 [Wn,pos] = min(Wn_theta);
24 theta_n = a + (pos-1)*PAS;

```

### A.4.2 Exécution du test basé sur $W_n$

```
1 function [Wn,Rn,theta_n,PV,CV,test] = GoF_Wn_elliptical(Y,law,M,a,b,PAS)
2
3 [Wn,Rn,theta_n] = Wn_elliptical(Y,law,a,b,PAS);
4
5 %Parametric bootstrap
6 Wn_boot = zeros(M,1);
7 temp = size(Y); n = temp(1);
8 n = min(100,n);
9 parfor h=1:M
10     Y_boot = sim_elliptical(Rn,[law,theta_n],n);
11     [Wn_boot(h),~,~] = Wn_elliptical(Y_boot,law,a,b,PAS);
12 end
13
14 %Estimated critical value
15 CV = prctile(Wn_boot,95);
16
17 %Estimated p-value
18 PV = sum(Wn_boot>Wn) / M;
19
20 test = 0;
21 if (PV < 0.05)
22     test = 1;
23 end
```

## A.5 Adéquation à une loi univariée

### A.5.1 Fonction de répartition de quelques lois

```

1 function value = cdf_univ_elliptical(x,law,beta)
2
3 %Gamma : law = 1 ;
4 %Fisher : law = 2 ;
5 %Weibull : law = 3 ;
6 %Rayleigh : law = 4 ;
7 %Log-Normal : law = 5
8
9 if (law==1)
10     value = gamcdf(x,beta(1),beta(2));
11 elseif (law==2)
12     value = fcdof(x,beta(1),beta(2));
13 elseif (law==3)
14     value = wblcdf(x,beta(1),beta(2));
15 elseif (law==4)
16     value = raylcdf(x,beta(1));
17 elseif (law==5)
18     value = logncdf(x,beta(1),beta(2));
19 end

```

### A.5.2 Calcul de la statistique de test $S_n$

```

1 function Sn = Sn_elliptical(X,law,beta)
2
3 %Gamma : law = 1 ;
4 %Fisher : law = 2 ;
5 %Weibull : law = 3 ;
6 %Rayleigh : law = 4 ;
7 %Log-Normal : law = 5
8
9 temp = size(X); n = temp(1);
10
11 %Calcul de la statistique de test
12 A1 = 0; A2 = 0;
13 for i=1:n
14     A2 = A2 + ( 1 - (cdf_univ_elliptical(X(i),law,beta))^2 );
15     for j=1:n
16         A1 = A1 + (1 - cdf_univ_elliptical(max(X(i),X(j)),law,beta));
17     end
18 end
19 Sn = (A1/n) - A2 + (n/3);

```

### A.5.3 Exécution du test basé sur $S_n$

```

1 function [Sn,beta_n,PV,CV] = GoF_Sn_elliptical(X,law,G1,G2,M)
2
3 n = size(X); n = n(1);
4 s1 = size(G1); A = s1(2);
5 s2 = size(G2); B = s2(2);
6
7 %Test statistic
8 Sn_beta = zeros(A,B);
9 for i=1:A
10     for j=1:B
11         Sn_beta(i,j) = Sn_elliptical(X,law,[G1(i),G2(j)]);
12     end
13 end
14 [Sn,J] = min(min(Sn_beta));
15 [~,pos] = min(Sn_beta);
16 I = pos(J);
17 beta_n = [G1(I),G2(J)];
18
19 %Parametric bootstrap
20 n = min(n,100);
21 Sn_boot = zeros(M,1);
22 parfor h=1:M
23     Sn_beta_boot = zeros(A,B);
24     Xboot = sim_univ_elliptical(law,beta_n,n);
25     g1=G1; g2=G2;
26     for i=1:A
27         for j=1:B
28             Sn_beta_boot(i,j) = Sn_elliptical(Xboot,law,[g1(i),g2(j)]);
29         end
30     end
31     [Sn_boot(h),~] = min(min(Sn_beta_boot));
32 end
33
34 % Estimated critical value
35 CV = prctile(Sn_boot,95);
36
37 % Estimated p-value
38 PV = sum(Sn_boot>Sn) / M;

```

## A.6 Approximation de la puissance des tests

### A.6.1 Test basé sur $V_n$

```

1 function [Vn, test] = GoF_Vn_TableCV(Y, model)
2
3 temp = size(Y); d = temp(1);
4
5 % Test statistic and sample correlation matrix
6 [Vn, Rn] = Vn_elliptical(Y, model);
7
8 % Mean value of the off-diagonal elements of Rn
9 rho = (sum(sum(Rn)) - d)/(d*(d-1));
10
11 % Loading the table that corresponds to d
12 if (d==2)
13     TableCV = 'TableCV_Vn_d2';
14 elseif (d==3)
15     TableCV = 'TableCV_Vn_d3';
16 elseif (d==4)
17     TableCV = 'TableCV_Vn_d4';
18 end
19
20 % Line in 'TableCV' that corresponds to rho
21 temp = abs(TableCV(:,1) - rho);
22 [~, line] = min(temp);
23
24 % Column in 'TableCV' that corresponds to theta
25 law = model(1);
26 theta = model(2);
27
28 if (law==1)
29     CV = TableCV(pos, 2);
30 elseif (law==2)
31     if (theta==1)
32         CV = TableCV(line, 3);
33     elseif (theta==3)
34         CV = TableCV(line, 4);
35     elseif (theta==6)
36         CV = TableCV(line, 5);
37     elseif (theta==9)
38         CV = TableCV(line, 6);
39     end
40 elseif (law==3)
41     if (theta==1)
42         CV = TableCV(line, 7);
43     elseif (theta==2)
44         CV = TableCV(line, 8);
45     elseif (theta==5)
46         CV = TableCV(line, 9);
47     end

```



```

48 end
49
50 % Goodness-of-fit test
51 test = 0;
52 if (Vn > CV)
53     test = 1;
54 end

1 function P = Power_Vn(n, IT, d, modelH1, rho, modelHo)
2
3 % modelH1 : it is a 1 x 2 vector for the alternative model
4 % rho : it is the value of rho in R (equicorrelated)
5 % modelHo : it is a 1 x 2 vector for the model to be tested
6
7 R = rho*ones(d,d);
8 for i=1:d
9     R(i,i) = 1;
10 end
11
12 test = zeros(1, IT);
13 parfor i=1:IT
14     Y = sim_elliptical(R, modelH1, n);
15     [~, test(i)] = GoF_Vn_TableCV(Y, modelHo);
16 end
17
18 P = sum(test) / IT;

1 function POWER = Table_POWER_Vn(n, d, IT, modelH1, rho, modelHo)
2
3 % modelH1 : it is a 1 x 2 matrix containing the I alternative models
4 % rho : it is a 1 x J vector containing the values of rho in R (equicorrelated)
5 % modelHo : it is a K x 2 matrix containing the K models to be tested
6 Temp = size(modelH1); I = Temp(1);
7 Temp = size(rho); J = Temp(2);
8 Temp = size(modelHo); K = Temp(1);
9
10 POWER = zeros(1, J, K);
11
12 for i=1:I
13     for j=1:J
14         for k=1:K
15             POWER(i, j, k) = Power_Vn(n, IT, d, modelH1(i,:), rho(j), modelHo(k,:));
16         end
17     end
18 end

1 function Table = Table_CV_Vn(n, IT, model, d)
2
3 Table = zeros(20, 1);

```

```

4 mu = zeros(1,d);
5
6 for t=1:20
7     theta = (t-1)/20;
8     Sigma = theta*ones(d,d);
9     for i=1:d
10        Sigma(i,i)=1;
11    end
12
13    Sn = zeros(1,IT);
14    parfor k=1:IT
15        if (model(1) == 1)
16            Y = mvnrnd(mu,Sigma,n);
17        elseif (model(1) == 2)
18            Y = mvtrnd(Sigma,model(2),n);
19        elseif (model(1) == 3)
20            Y = mvpearsonII(Sigma,model(2),n);
21        end
22
23        [R,~] = Elliptique_R(Y,model,0);
24        B1 = 0; B2 = 0;
25        for i=1:n
26            B2 = B2 + ( 1 - (cdf_R(R(i),[model,d]))^2 );
27            for j=1:n
28                B1 = B1 + (1 - cdf_R(max(R(i),R(j)),[model,d]));
29            end
30        end
31        Sn(k) = (B1/n) - B2 + (n/3);
32    end
33
34    Table(t) = prctile(Sn,95);
35    clear Sn;
36 end

```

A.6.2 Test basé sur  $W_n$ 

```

1  function [Wn, test , TableCV] = GoF_Wn_TableCV(Y, law)
2
3  temp = size(Y); d = temp(2);
4
5  if (law==2)
6      a = 1; b = 10; PAS = 1;
7  elseif (law==3)
8      a = -.5; b = 9.5; PAS = 1;
9  end
10
11 % Test statistic, sample correlation matrix, and theta_n
12 [Wn,Rn,theta_n] = Wn_elliptical(Y,law,a,b,PAS);
13
14 % Mean value of the off-diagonal elements of Rn
15 rho = (sum(sum(Rn)) - d)/(d*(d-1));
16
17 % Loading the table that corresponds to d
18 if (d==2)
19     TableCV = load('TableCV_Wn_d2.mat');
20 elseif (d==3)
21     TableCV = load('TableCV_Wn_d3.mat');
22 end
23
24
25 % Line in 'TableCV' that corresponds to rho
26 temp = abs(TableCV.X(:,1) - rho);
27 [~,line] = min(temp);
28
29 %Column in 'TableCV' that corresponds to theta
30
31 if (law==2)
32     if (theta_n <= 2)
33         CV = TableCV.X(line,2);
34     elseif (theta_n <= 4.5)
35         CV = TableCV.X(line,3);
36     elseif (theta_n <= 7.5)
37         CV = TableCV.X(line,4);
38     else
39         CV = TableCV.X(line,5);
40     end
41 elseif (law==3)
42     if (theta_n <= 1.5)
43         CV = TableCV.X(line,6);
44     elseif (theta_n <= 3.5)
45         CV = TableCV.X(line,7);
46     else
47         CV = TableCV.X(line,8);
48     end
49 end
50

```

```

51 % Goodness-of-fit test
52 test = 0;
53 if (Wn > CV)
54     test = 1;
55 end

1 function P = Power_Wn(n, IT, d, modelH1, rho, lawHo)
2
3 % modelH1 : it is a 1 x 2 vector for the alternative model
4 % rho : it is the value of rho in R (equicorrelated)
5 % lawHo : it is a number for the elliptical law to be tested
6
7 R = rho*ones(d,d);
8 for i=1:d
9     R(i,i) = 1;
10 end
11
12 test = zeros(1, IT);
13 parfor i=1:IT
14     Y = sim_elliptical(R, modelH1, n);
15     [~, test(i)] = GoF_Wn_TableCV(Y, lawHo);
16 end
17
18 P = sum(test) / IT;

1 function POWER = Table_POWER_Wn(n, d, IT, modelH1, rho, lawHo)
2
3 % modelH1 : it is a 1 x 2 matrix containing the I alternative models
4 % rho : it is a 1 x J vector containing the values of rho in R (equicorrelated)
5 % lawHo : it is a number for the elliptical law to be tested
6 Temp = size(modelH1); I = Temp(1);
7 Temp = size(rho); J = Temp(2);
8
9 POWER = zeros(I, J);
10
11 for i=1:I
12     for j=1:J
13         POWER(i, j) = Power_Wn(n, IT, d, modelH1(i, :), rho(j), lawHo);
14     end
15 end

1 function Table = TableCV_Wn(n, IT, law, d, rho, theta)
2
3 PAS = .25;
4
5 temp = size(rho); I = temp(2);
6 temp = size(theta); J = temp(2);
7 Table = zeros(I, J);
8

```

```
9  for i=1:I
10     R = rho(i)*ones(d,d);
11     for k=1:d
12         R(k,k) = 1;
13     end
14
15     for j=1:J
16         model = [law,theta(j)];
17         t = theta(j);
18         parfor h=1:IT
19             Y = sim_elliptical(R,model,n);
20             [Wn(h),~,~] = Wn_elliptical(Y,law,t-2,t+2,PAS);
21         end
22         Table(i,j) = prctile(Wn,95);
23     end
24 end
```